

Hálózati folyamatok modellezése differenciálegyenletekkel

Doktori értekezés

Nagy Noémi

ELTE Matematika Doktori Iskola

Dr. Laczkovich Miklós

Alkalmazott matematika program

Michaletzky György

Témavezető:

Simon L. Péter

egyetemi tanár, az MTA doktora



Alkalmazott Analízis és Számításmatematikai Tanszék

Eötvös Loránd Tudományegyetem

Természettudományi Kar

2015

Tartalomjegyzék

1. Bevezetés	1
2. Az általános matematikai modell	6
2.1. A hálózat és a dinamika meghatározása	6
2.2. Az alapegyenletek felírása	7
2.3. Állapottér 2 lehetséges csúcsállapot esetén	9
2.4. A klasszikus <i>SIS</i> modell	10
2.4.1. A klasszikus modell bemutatása 3 csúcsú teljes gráf esetén	11
2.4.2. Alacsony dimenziós modellek	14
2.5. A numerikus szimuláció algoritmus	18
3. Járványterjedés modellezése véletlen gráfokkal megadott hálózatokon	21
3.1. Motiváció	21
3.2. A modell formalizálása	23
3.2.1. A modell megvalósíthatósága	26
3.2.2. A fertőződésért felelős átmeneti ráták meghatározása	28
3.3. Elméleti becslés az <i>II</i> élek várható értékére	29
3.3.1. A reguláris véletlen gráf esete	30
3.3.2. Tetszőleges fokszámeloszlású gráf esete	32
3.4. Elméleti becslés az <i>SI</i> élek várható értékére	35
3.4.1. Explicit formula az $e_{SI}^{\infty}(k)$ értékekre reguláris véletlen gráf esetén	37
3.4.2. Az <i>SI</i> élek maximális száma τ függvényében	39
3.5. Az eljárás áttekintése	41
3.6. Összefoglalás	42

4. Betegségterjedés modellezése speciális struktúrájú gráf esetén	45
4.1. Motiváció	45
4.2. Redukált alapegyenletek a betegszám függvényében	46
4.2.1. Becslés az SI élek átlagos számára az állapotok súlyozásával	47
4.2.2. Becslés az SI élek átlagos számára lineáris közelítéssel	53
4.3. Alapegyenletek a front hosszára	55
4.4. Az élek szintjén felírt egyenletek	59
4.5. Összefoglalás	61
5. Alacsony dimenziós közelítő rendszerek analízise	64
5.1. Motiváció	64
5.2. Átlagoláson alapuló közelítő modellek	65
5.2.1. Az átlagoláson alapuló egyenlet	65
5.2.2. A párok szintjén felírt egyenletek	66
5.3. Közelítő modellek heterogén fokszámeloszlású hálózatokon	68
5.3.1. Fokszám szerinti átlagoláson alapuló és párok szintjén felírt egyenletek	68
5.3.2. Párok szintjén felírt kompakt egyenletrendszer	70
5.4. Összefoglalás	76
6. Hálózati modell együttes csődvalószínűségek meghatározására	79
6.1. Motiváció	79
6.2. A probléma felvázolása és a Gauss kopulás módszer bemutatása	80
6.3. A hálózati megközelítésen alapuló modell	82
6.3.1. A hálózati modell leírása	82
6.3.2. Az eljárás ismertetése a csődvalószínűségek meghatározására	84
6.4. A csúcsok szintjén felírt egyenletek	86
6.4.1. A modell formalizálása	87
6.4.2. A modell vizsgálata	92
6.5. Összefoglalás	97

Köszönetnyilvánítás

Köszönettel tartozom elsősorban témavezetőmnek, Simon Péternek lelkiismeretes munkájáért, végtelen türelméért és megértő támogatásáért. A kutató munkánk során nyújtott teljes körű szakértelme, zseniális ötletei, inspiráló észrevételei és dicsérő szavai nagyban hozzájárultak a doktori értekezésem elkészüléséhez. Külön köszönöm Neki, hogy mindig rendelkezésemre állt, ha bármilyen problémám adódott, illetve azt is, hogy a jó hangulatú megbeszélésekkel örömtelivé tette számomra a munkafolyamatot.

Szeretném megköszönni Bihary Zsoltnak, hogy figyelmünkbe ajánlotta a pénzügyi piacokon megjelenő problémát, így lehetőséget adva egy új kutatási témakör elindítására, mely során hasznos tanácsokkal látott el minket.

Ezen kívül köszönöm az Alkalmazott Analízis és Számításmatematikai Tanszék dolgozóinak biztatását és támogatását, kiváltképp Faragó Istvánnak, aki lehetővé tette, hogy a tanszéken folyó kutatásoknak aktív részese legyek, illetve Havasi Ágnesnek, aki kiváló angol tudásával segítette a munkámat.

Hálával tartozom továbbá szüleimnek és testvéreimnek, akik nélkül ez a munka nem jöhetett volna létre. Köszönöm nekik, hogy tanulmányaim során türelemmel és megértéssel támogattak, és minden helyzetben mellettem álltak. Külön köszönet Édesanyámnak, aki a dolgozat végső átnézésében is részt vett.

1. fejezet

Bevezetés

A hálózatok és a hálózatokon zajló folyamatok vizsgálata jelentős tudományterületté fejlődött az utóbbi fél évszázadban, melyről több remek összefoglaló mű is született [8, 12, 21]. Hálózatokkal az élet számos területén találkozhatunk, ami jól magyarázza, hogy ez a kutatási téma miért került a figyelem középpontjába. Példaként említhető valós hálózatokra többek között a neuron hálózatokon terjedő folyamatok, az epidemiológiai folyamatok, a vélemény terjedés szociális hálózatokon vagy az interneten, illetve az üzleti életben felmerülő hálózatokkal leírható folyamatok [8, 21]. A hálózatok segítségével hatékonyan reprezentálható az egymással kölcsönhatásban álló elemek viselkedése, legyen szó akár személyekről, társadalmi csoportokról vagy városokról. A hálózatokon vizsgált folyamatok viselkedésének megértésével lehetőség nyílik valós problémák megoldására, például járványok megfékezésére vagy gazdasági válságok kezelésére. Kezdetben a hálózatok fejlődését és növekedését tanulmányozták, később a hálózatokon végbemenő dinamikai folyamatok elemzése egy külön kutatási területté fejlődött.

Attól függően, hogy milyen hálózati struktúrát követelünk meg, illetve hogyan változtatjuk a folyamat dinamikáját, számos modellel találkozhatunk. Ezek közös jellemzője az a struktúra amiben ezeket vizsgálják, az N csúcsú gráf, amin zajlik a sztochasztikus folyamat, és a dinamika, amely meghatározza hogy a csúcsok hányféle állapotban lehetnek, illetve az állapotok közötti átmenetek szabályait is rögzíti. A valós hálózatok meglehetősen komplex, bonyolult rendszerek, így a kutatások során leegyszerűsített szerkezetű modelleken dolgoznak, emellett a dinamikát is úgy adják meg, hogy a hálózati modell minél könnyebben kezelhető legyen, de azért néhány, számunkra fontos jellemzőjét megőrizze valamilyen formában.

Jelentősebb hálózat típusok áttekintése

Rengeteg módszer ismert a hálózati struktúra modellezésére, melyekről bővebb információt a [10, 11, 22] könyvekben találhatunk. A továbbiakban csak néhány történeti, illetve matematikai szempontból fontos hálózati modellt ismertetek. Az egyik első modell Erdős Pál és Rényi Alfréd nevéhez fűződik [11]. Az ő megközelítésük szerint, az életben megfigyelhető hálózatok a véletlen szerveződési elvén alakulnak ki, ezért az általuk ajánlott modell esetén az Erdős-Rényi gráf csúcsai között p valószínűséggel alakul ki él, ahol $p \in (0, 1)$ rögzített paraméter. Ennek a gráfnak a $p(n)$ fokszámeloszlása binomiális eloszlást követ, ahol $p(n)$ jelöli a n fokú csúcsok arányát az összes csúcshoz képest. A valós hálózatokat vizsgálva jötték rá arra, hogy ezen hálózatok rendelkeznek a "kis-világ" tulajdonsággal, ami azt jelenti, hogy az átlagos úthossz két csúcs között sokkal kisebb, mint a csúcsok száma, nevezetesen egy N csúcsú gráfon az átlagos úthossz $\log N$ nagyságrendű. Az Erdős-Rényi gráf $p > 1/N$ esetén rendelkezik a "kis-világ" tulajdonsággal, azonban az életben létrejövő hálózatokra jellemző klaszterezettségi szintet nem éri el. Egy gráf klaszterezettségi együtthatója azt méri, hogy tetszőleges csúcs két szomszédja milyen valószínűséggel van összekötve, tehát ez a mennyiség a háromszögek arányát adja meg az összes kapcsolódó élpárhoz képest. Felmerült az igény az Erdős-Rényi gráf klaszterezettségi együtthatójánál nagyobb klaszterezettségű gráfok használatára, amik jobban modellezik az életben megfigyelhető hálózatokat. Így jött létre a Watts-Strogatz modell [27], ahol a gráf létrehozásánál kiindulnak egy rácsból (például egy körgráfból), majd véletlenszerűen behúznak éleket a távoli csúcsok között, ezzel növelve a klaszterezettségi szintet a kívánt mértékig. Ebben a modellben - megfelelő paraméterek mellett - teljesül a "kis-világ" tulajdonság is, és a $p(n)$ értékek exponenciálisan csökkennek. A fokszámeloszlás ezen tulajdonsága azonban egyáltalán nem jellemző a valós hálózatokra, melyek tanulmányozása során azt vették észre, hogy nagy n esetén a fokszámeloszlás hatványfüggvény alakban adható meg, azaz $p(n) \approx cn^{-\alpha}$, ahol c konstans és α általában egy 2 és 3 közötti paraméter. Ez az jelenti, hogy habár a valós hálózatok esetén a legtöbb csúcs foka kicsi és nagy fokú csúcsok csak kis valószínűséggel fordulnak elő a gráfban, de több magas fokú csúcs van, mint az Erdős-Rényi gráfban. Ezek fontos jelentőségűek a gráfon vizsgált folyamatok lezajlása szempontjából, ezért középpontoknak is nevezik őket. Barabási Albert-László és Albert Réka kutatásai eredményeként lett közzismert ez a tény, és ők ajánlottak egy eljárást, ami ilyen fokszámeloszlású véletlen gráfot generál [7]. Az algoritmusuk a következő: vegyünk egy kis méretű tetszőleges gráfot, majd egyenként vegyünk hozzá előre

meghatározott, azonos fokú csúcsokat, míg el nem érjük az N csúcsú hálózatot. Minden új csúcsot úgy kössünk be a már meglévő gráfba, hogy a nagyobb fokszámú csúcsokhoz nagyobb valószínűséggel csatlakozzanak. Ezt úgy lehet röviden megfogalmazni, hogy a gazdag, egyre gazdagabb lesz, azaz - a valós életben létrejövő hálózatoknak megfelelően - egy sok ismerőssel rendelkező ember, több új kapcsolatot tud létrehozni, mint egy kevés kapcsolattal rendelkező. Az így generált hálózatokat hívjuk preferált kapcsolódással létrehozott véletlen gráf modellnek. Ez a hálózat tehát rendelkezik a valós hálózatokra jellemző fokszámeloszlással, emellett a "kis-világ" tulajdonsággal is, azonban a klaszterezettségi szintje nem éri el a valós hálózatokban mért értéket.

A valódi hálózatoktól eltérő, de matematikailag fontos véletlen gráftípus a Bollobás-féle konfigurációs modellel generált hálózat [9], melyre a továbbiakban a konfigurációs gráf elnevezést is használni fogjuk. A Bollobás-féle eljárással egy előre megadott fokszámeloszlású gráfot lehet előállítani. A létrehozandó gráfot a következő módon adjuk meg: legyen N_l az n_l fokú csúcsok száma az N csúcsú gráfban, azaz a fokszámeloszlás: $p(n_l) = N_l/N$, $l = 1, 2, \dots, L$ és L jelölje a különböző fokok számát. Az algoritmus során minden egyes csúcsból kiindítunk annyi fél-élet, amennyi hozzá tartozik, tehát N_l darab csúcsból n_l darab fél-él fog kiállni, $l = 1, \dots, L$. Ezeket a fél-éleket kötjük össze véletlenszerűen, tekintettel arra, hogy többszörös- és hurokél ne legyen a hálózatban. A konfigurációs modellel generált gráfok egy fontos típusa a reguláris véletlen gráf, ahol minden csúcs foka megegyezik. Emellett heterogén fokszámeloszlású gráfra gyakran használt példa a bimodális gráf, ahol $L = 2$, tehát a lehetséges fokok száma kettő, és a csúcsok fele n_1 fokú, míg a másik fele n_2 fokú. A bimodális gráf is könnyedén előállítható a konfigurációs modell segítségével, de tetszőleges fokszámeloszlást is meg lehet adni, csak ki kell számítani az adott fokszámokhoz az N_l értékeket az N csúcshoz tartozó gráf esetén. A Bollobás-féle módszerrel ugyanolyan gráfot kapunk eredményül, mintha tekintenénk az összes többszörös- és hurokélmentes hálózatot az előre megadott fokszámeloszlás mellett, majd ezen halmazból kiválasztanánk egy gráfot véletlenszerűen (egyenletes eloszlással).

Megjegyezzük, hogy az ilyen módon előállított, azaz egy előre meghatározott előállítású vagy fokszámeloszlású gráfokat tartalmazó osztályból egyenletes eloszlás szerint kiválasztott gráfokra a véletlen gráf elnevezéssel is szoktunk hivatkozni, feltéve hogy a megadott gráf-halmazban megtalálható az ilyen gráfok összes lehetséges megvalósulása. Ezen meghatározás

szerint az Erdős-Rényi gráf, a Watts-Strogatz gráf, a Barabási-Albert gráf és a Bollobás-féle konfigurációs modellel előállított gráfok is véletlen gráfoknak tekinthetők.

Ezek a gráftípusokon kívül érdemes még megemlíteni a speciális struktúrájú hálózatokat, melyek esetén a gráf egyedi felépítéséből adódóan van esély az általános modelleknél hatékonyabb eljárásokat kidolgozni a rajtuk végbemenő folyamatok elemzésére. Ilyen speciális struktúrájú hálózat például a teljes gráf, a körgráf, a csillag gráf, vagy a háztartás típusú modellek [6, 15].

Járványterjedés kompartment modelljei

A kutatás során legfontosabb hálózat típusok áttekintése után néhány alapvető kompartment modellt ismertetek a hálózaton megadott dinamika leírására, melyekről bővebben a [5, 13] könyvekben olvashatunk. Az értekezésben minden problémára járványterjedési folyamatként fogunk tekinteni, azért csak az ezeket meghatározó dinamikákról teszünk említést. Az egyik leggyakrabban használt kompartment modell egy S -egészséges (susceptible) és egy I -beteg (infected) csoportból áll. Az egészséges egyedeket megfertőzhetik a betegek, amennyiben kapcsolatban állnak, illetve a betegek meggyógyulhatnak, azaz $S \rightarrow I$ és $I \rightarrow S$ átmenet is lehetséges, tehát itt egy ciklikus folyamatról beszélhetünk. Ezt hívjuk SIS típusú járványterjedési modellnek. Ehhez képest a SIR -kompartment modell egyirányú folyamatot határoz meg. Ebben az esetben három osztályt különböztetünk meg: S -egészséges, I -beteg és R -immunis (recovered) és a lehetséges átmenetek: $S \rightarrow I$, $I \rightarrow R$. Ez azt jelenti, hogy az egészségeseket megbetegíthetik a betegek, majd a betegek meggyógyulás után immunissá válnak a betegségre, és ebből az állapotból már nem léphetnek ki. Ismertek még ezeknek a járványterjedést leíró dinamikáknak különböző változatai, a teljesség igénye nélkül ismertetünk néhányat: az SIS dinamika egyirányúvá tehető, ilyenkor csak $S \rightarrow I$ átmenet fordulhat elő, ezt hívjuk SI dinamikának. Az SIR dinamikát többféleképpen is variálhatjuk: $SIRS$ dinamika esetén az immunis egyed újra egészségesé válhat, így a folyamat periodikussá válik; az $SEIR$ dinamika során, amennyiben a betegségnek van egy lappangási időszaka, beépíthető egy E -lappangó csoport (exposed), így megjelennek az $S \rightarrow E$ és az $E \rightarrow I$ átmenetek az eddigiek mellett, de a folyamat egyirányú marad.

Az értekezés szerkezete

A bevezető lezárásaként az értekezés felépítéséről írunk. A doktori kutatásaim során *SIS* és *SI* típusú járványterjedési folyamatokat vizsgáltunk sokféle gráftípuson. A matematikai modell, ami leírja ezeket a folyamatokat egy folytonos paraméterű Markov-lánc, M^N elemszámú állapottérrel, amennyiben a csúcsok M -féle lehetséges állapotban lehetnek. Az ehhez kapcsolódó - a folyamatot pontosan megadó - un. alapegyenletek egy M^N egyenletből álló lineáris differenciálegyenlet-rendszert alkotnak, melyeket numerikusan megoldani az általában vizsgált nagy csúcsszámú hálózatok esetén már nem lehet. Az értekezés célkitűzése egyrészt az alapegyenletek számának csökkentése különböző technikákkal, másrészt ismert járványterjedési modellek elemzése, harmadrészt a hálózati modellek alkalmazásának kiterjesztése új problémák megoldására.

A 2. fejezet témája a matematikai modell pontos felépítése és alapvető alacsony dimenziós módszerek illetve a numerikus szimuláció ismertetése. Az ezt követő négy fejezetben saját kutatási eredményeket dolgozunk fel. A fejezetek a járványterjedés modellezési lehetőségeit vizsgálják különböző esetekben, a témakörök gondolatai sok ponton összekapcsolódnak, de mindegyik fejezet önálló munkaként is megállja a helyét.

A 3. és a 4. fejezetekben az alapegyenletek számának a csökkentésére ajánlunk különböző eljárásokat. A kérdéskörök legfontosabb motivációja az átlagos betegszám meghatározására az idő függvényében, amennyiben a járvány szétterjed a gráfon, tehát nem foglalkozunk olyan esettel, amikor a betegség megszűnik. A módszerek hatékonyságának igazolására az eredményeket a numerikus szimuláció megoldásával vetettük össze. A 3. fejezetben különböző véletlen gráfok szolgáltatják a hálózatot [1], amin a vizsgálatokat végezzük, míg a 4. fejezetben speciális struktúrájú gráfon dolgozunk [2]. Az 5. fejezetben először bemutatjuk néhány ismert, alacsony dimenziós közelítő modell analitikus eredményeit, majd hasonló állításokat fogalmazunk meg egy - az előzőekhez kapcsolódó - újabb rendszer esetén. A 6. fejezetben egy hálózati modellt kínálunk egy pénzügyi probléma leírására, majd a modellt átalakítva egy numerikusan jól kezelhető eljárást adunk a cégek együttműködése és csődbe kerülése folytán megjelenő fontos mennyiség, a rendszerkockázat mérésére. Az utolsó két fejezet eredményéről beszámoló tudományos publikációk előkészületben vannak [3, 4].

Az értekezésben alkalmazott numerikus számításokat és szimulációkat saját készítésű Matlab programok segítségével végeztük el.

2. fejezet

Az általános matematikai modell

Ebben a fejezetben bemutatjuk a pontos matematikai modellt, ami alkalmas a betegségterjedés folyamatának leírására. Ennek az általános modellnek néhány speciális esetét vizsgáltuk meg a doktori kutatás során.

2.1. A hálózat és a dinamika meghatározása

A hálózat leírásához vegyünk egy N csúcú gráfot, melyet az adjacencia mátrixával adunk meg: $G = ((g_{p,q}))_{p,q=1,\dots,N}$, ahol $g_{p,p} = 0$, $p = 1, \dots, N$. A legegyszerűbb esetben a G mátrixban tároljuk, hogy mely csúcsok között vezet él: $g_{p,q} = 1$, ha az p -edik és q -edik csúcs között fut él, ha pedig nincs közöttük él, akkor $g_{p,q} = 0$, ilyenkor fel fogjuk tenni, hogy többszörös- és hurokélmentes a hálózat, azaz egy egyszerű gráfot tekintünk. Általánosabb esetben a G mátrix elemei tetszőleges valós számok is lehetnek, amiket élsúlyoknak nevezünk és a $g_{p,q}$ eleme a mátrixnak kifejezi, hogy milyen erősséggel hat a p -edik csúcs állapota a q -edik csúcs állapotváltozására. Szimmetrikus mátrix esetén irányítatlan gráfról beszélünk.

A dinamika leírására legyen adott egy véges állapot halmaz $\{m_1, m_2, \dots, m_M\}$, ami megadja az egyes csúcsok M darab különböző lehetséges állapotát. Ezek alapján az állapottér M^N elemű. A továbbiakban az állapottérre azonosítsuk az $\mathcal{S} := \{\underline{a} = (a_1, \dots, a_N) : a_n \in \{m_1, m_2, \dots, m_M\}, n = 1, \dots, N\}$ halmazzal. Az állapottér egy $\underline{a} = (a_1, \dots, a_N) \in \mathcal{S}$ eleméből pontosan akkor juthatunk el közvetlenül a $\underline{b} = (b_1, \dots, b_N) \in \mathcal{S}$ elembe, ha létezik egyetlen $n \in \{1, 2, \dots, N\}$, melyre $a_n \neq b_n$ és $a_l = b_l$, ha $l \neq n$, $l = 1, \dots, N$. Tehát a gráf két állapota között akkor létezik közvetlen átmenet, ha pontosan egy csúcs állapotban térnek el egymástól,

és ilyenkor ezen csúcs állapotában következik be állapotváltozás. Egy csúcs állapotváltozását egy exponenciális eloszlású valószínűségi változó írja le, azaz $1 - e^{-f_{ij}\delta t}$ legyen a valószínűsége annak, hogy $\delta t > 0$ idő alatt bekövetkezik az $m_i \rightarrow m_j$ állapotváltozás, ahol f_{ij} az exponenciális eloszlás paraméterét adja meg. Ezt az f_{ij} valós számot az állapotváltozás rátájának hívjuk. Egy csúcs m_i állapotból m_j állapotba jutása függ a G adjacencia mátrixtól, a folyamat aktuális állapotától és attól is, hogy melyik csúcs állapotváltozását vizsgáljuk. Tehát az \mathcal{S} állapottéren bekövetkező

$$(a_1, \dots, a_{n-1}, a_n, a_{n+1}, \dots, a_N) \rightarrow (a_1, \dots, a_{n-1}, b_n, a_{n+1}, \dots, a_N)$$

közvetlen állapotváltozás rátáját az $f_{ij} = f_{ij}(n, G, \underline{a})$ függvény adja meg, amennyiben $a_n = m_i$ illetve $b_n = m_j$ és az f_{ij} függvény változói: n , ugyanis a n -edik csúcs állapota változik meg, G az adjacencia mátrix és $\underline{a} \in \mathcal{S}$ a folyamat jelenlegi állapota. Az állapothalmaz és az átmenetek rátáit megadó függvények segítségével meghatároztuk a dinamikát.

2.2. Az alapegyenletek felírása

Mivel az állapottéren a közvetlen átmenetek valószínűségeit exponenciális eloszlás írja le, így a folyamatot egy folytonos paraméterű M^N állapotterű X Markov-láncnak tekinthetjük. $X(t)$ jelölje azt a valószínűségi változót, ami megadja, hogy milyen állapotban van a sztochasztikus folyamat a t időpontban, $t \in [0, +\infty)$. A folyamathoz kapcsolódóan felírhatóak az ún. alapegyenletek, azaz egy állandó együtthatós lineáris differenciálegyenlet-rendszer, aminek az $x_{\underline{a}}(t)$ változói megadják, hogy mekkora valószínűséggel van a Markov-folyamat az \underline{a} állapotban a t időpontban, tehát $x_{\underline{a}}(t) = P(X(t) = \underline{a})$, minden $\underline{a} \in \mathcal{S}$ esetén.

Most levezetjük, hogyan lehet felírni az alapegyenleteket. Vegyük az \mathcal{S} állapottér két tetszőleges elemét: $\underline{a} = (a_1, \dots, a_N)$ és $\underline{b} = (b_1, \dots, b_N)$. Vizsgáljuk meg, hogy mi a valószínűsége, hogy δt idő alatt bekövetkezik az állapottéren az $\underline{a} \rightarrow \underline{b}$ átmenet, amennyiben az \underline{a} és \underline{b} állapot pontosan egy csúcs állapotában tér el egymástól:

$$P(X(t + \delta t) = \underline{b} | X(t) = \underline{a}) = 1 - e^{-f_{ab}\delta t},$$

ahol f_{ab} -vel jelöljük az $\underline{a} \rightarrow \underline{b}$ közvetlen állapotváltozáshoz tartozó rátát. Az exponenciális függvény Taylor-sora alapján ismert, hogy $1 - e^{-f_{ab}\delta t} = f_{ab}\delta t + o(\delta t)$, ahol $\frac{o(\delta t)}{\delta t} \rightarrow 0$, ha

$\delta t \rightarrow 0$. Tehát

$$P(X(t + \delta t) = \underline{b} | X(t) = \underline{a}) = f_{\underline{a}\underline{b}}\delta t + o(\delta t).$$

Abban az esetben, ha az \underline{a} és \underline{b} állapotok legalább kettő csúcs állapotában térnek el, akkor $P(X(t + \delta t) = \underline{b} | X(t) = \underline{a}) = o(\delta t)$, az egymás után bekövetkező közvetlen állapotváltozások függetlenségére tekintettel.

Amennyiben $\underline{a} = \underline{b}$, azaz nem történik δt idő alatt állapotváltozás, akkor

$$P(X(t + \delta t) = \underline{b} | X(t) = \underline{b}) = (1 - f_{\underline{b}*}\delta t) + o(\delta t),$$

ahol $f_{\underline{b}*}$ jelöli azon ráták összegét, amik alapján a folyamat a \underline{b} állapotból átkerül bármelyik másik, egy csúcs állapotában különböző állapotba.

Tekintsük az \mathcal{S} állapottér egy tetszőleges \underline{b} elemét. Az állapottér bármely állapotából elérhető a \underline{b} állapot δt idő alatt valamilyen valószínűséggel. Alkalmazva a teljes valószínűség tételét:

$$x_{\underline{b}}(t + \delta t) = P(X(t + \delta t) = \underline{b}) = \sum_{\underline{a} \in \mathcal{S}} P(X(t + \delta t) = \underline{b} | X(t) = \underline{a}) P(X(t) = \underline{a}) =$$

$$\sum_{\underline{a} \in \mathcal{S}_{\underline{b}}, \underline{a} \neq \underline{b}} f_{\underline{a}\underline{b}}\delta t x_{\underline{a}}(t) + (1 - f_{\underline{b}*}\delta t)x_{\underline{b}}(t) + o(\delta t),$$

ahol $\mathcal{S}_{\underline{b}}$ jelöli az \mathcal{S} állapottér azon elemeit tartalmazó halmazt, melyek a \underline{b} állapottól pontosan egy csúcs állapotában térnek el. Ezt a jelölést bevezetve $f_{\underline{b}*}$ is megadható képlettel: $f_{\underline{b}*} = \sum_{\underline{a} \in \mathcal{S}_{\underline{b}}, \underline{a} \neq \underline{b}} f_{\underline{b}\underline{a}}$.

Az egyenlet mindkét oldalából kivonva $x_{\underline{b}}(t)$ -t, majd δt -vel osztva, határátmenetet véve $\delta t \rightarrow 0$ esetén azt kapjuk, hogy

$$\dot{x}_{\underline{b}}(t) = \sum_{\underline{a} \in \mathcal{S}_{\underline{b}}, \underline{a} \neq \underline{b}} f_{\underline{a}\underline{b}}x_{\underline{a}}(t) - f_{\underline{b}*}x_{\underline{b}}(t), \quad \underline{b} \in \mathcal{S}, \quad (2.1)$$

Ezeket az egyenleteket hívjuk alapegyenleteknek, azaz kaptunk egy M^N egyenletből álló állandó együtthatós lineáris differenciálegyenlet-rendszert, tömörebb formában:

$$\dot{\mathbf{x}}(t) = \mathbf{P}\mathbf{x}(t), \quad (2.2)$$

ahol $\mathbf{x}(t) = (\dots, x_{\underline{b}}(t), \dots)^T$ M^N dimenziós oszlopvektor, $\underline{b} \in \mathcal{S}$, és a $M^N \times M^N$ dimenziós \mathbf{P} mátrixban tároljuk a differenciálegyenlet-rendszer együtthatóit, pontosabban az átmeneti rátákat és a főátlóbeli elemek: $-f_{\underline{b}*}$, $\underline{b} \in \mathcal{S}$. Fontos megemlíteni, hogy a \mathbf{P} mátrix mindig egy

sztochasztikus mátrix lesz a rendszer zárttsága miatt, azaz minden oszlopban az elemek összege nulla.

Megjegyezzük, hogy azok az állapotok, melyek a \underline{b} állapottól több, mint egy csúcs állapotában eltérnek, nem jelennek meg a (2.1) egyenlet jobb oldalán. Így az alapegyenletek felírása konkrét állapottér és dinamika esetén nem annyira bonyolult, hiszen csak a közvetlen átmene-
teket meghatározó rátákra lesz szükség.

A (2.2) alapegyenletek a Markov-folyamat pontos modelljét adják, azonban nagy méretű hálózatok esetén a M^N méretű differenciálegyenlet-rendszer numerikusan kezelhetetlen.

2.3. Állapottér 2 lehetséges csúcsállapot esetén

Ebben a részben részletesebben megvizsgáljuk a Markov-folyamat állapotterének felépítését és az átmenetvalószínűségeket abban az esetben, ha bármely csúcs két állapotban lehet, tehát $M = 2$. A lehetséges állapotok halmaza: $\{m_1, m_2\}$ és az állapottér 2^N elemű. Ilyen dinamikák például az *SIS* illetve az *SI* dinamika, amiket a doktori kutatásom során is vizsgáltunk.

Ebben az esetben lehetőség van egy egyszerűbb formában vizsgálni a folyamat viselkedését. Ehhez bontsuk fel az $\mathcal{S} = \{m_1, m_2\}^N$ állapotteret $N + 1$ részhalmazzra a következőképpen: jelölje \mathcal{S}^k a k darab m_2 állapotú csúcsot tartalmazó állapotok részhalmazát, így \mathcal{S}^k elemei: $\{S_1^k, S_2^k, \dots, S_{d_k}^k\}$, ahol $d_k = \binom{N}{k}$, $k = 0, \dots, N$ és $S_i^k \in \mathcal{S}^k$ jelöli az i -edik elemét az \mathcal{S}^k halmaznak. Tehát például \mathcal{S}^0 csak az $S_1^0 = (m_1 m_1 \dots m_1)$ állapotot tartalmazza, azaz a gráfban ekkor minden csúcs m_1 állapotú, míg \mathcal{S}^N egyetlen eleme: $S_1^N = (m_2 m_2 \dots m_2)$. Az \mathcal{S}^k halmazok diszjunkt részhalmazokra bontják fel az \mathcal{S} állapotteret és $\mathcal{S} = \bigcup_{k=0}^N \mathcal{S}^k$.

Két állapot esetén csak az $m_1 \rightarrow m_2$ és az $m_2 \rightarrow m_1$ átmenetek rátáit megadó függvényeket kell meghatároznunk, vagyis az f_{12} és f_{21} függvényeket. Megjegyezzük, hogy az állapottér felbontása lehetővé teszi a közvetlen átmenetek pontosabb leírását. Egy csúcs $m_1 \rightarrow m_2$ átmenete esetén az állapottéren egy $S_i^k \rightarrow S_j^{k+1}$ típusú átmenet történik, ahol S_i^k és S_j^{k+1} olyan állapotok, amik egyetlen egy csúcs állapotában térnek csak el, azaz $\exists! n \in \{1, \dots, N\} : S_i^k(n) = m_1$, $S_j^{k+1}(n) = m_2$ és $S_i^k(l) = S_j^{k+1}(l)$, minden $l \neq n$ esetén, ahol $S_i^k(n)$ jelöli az állapottér S_i^k állapotában az n -edik csúcs állapotát. Ugyanígy az $m_2 \rightarrow m_1$ átmenet az állapottéren egy $S_i^{k+1} \rightarrow S_j^k$ típusú átmenetnek felel meg, ahol a S_i^{k+1} és S_j^k állapotok pontosan egy csúcs állapotában térnek el.

Ebben az esetben is felírhatjuk az alapegyenleteket, amelyek a szokásos $\dot{x}(t) = Px(t)$ lineáris differenciálegyenlet-rendszer alakot öltik, ahol az $x(t)$ oszlopvektorban tároljuk az egyes állapotok valószínűségeit a t időpontban. Legyen az x_k oszlopvektorban a k darab m_2 állapotú csúcsot tartalmazó összes lehetséges állapot valószínűsége, azaz

$$x_k := (x_{S_1^k}, x_{S_2^k}, \dots, x_{S_{d_k}^k})^T, \quad k = 1, \dots, N,$$

ahol $x_{S_i^k}$ jelöli a S_i^k állapot valószínűségét. Látható, hogy ekkor x előáll $(x_1^T, x_2^T, \dots, x_N^T)^T$ alakban. Ha 2 lehetséges állapotban lehetnek a csúcsok, akkor P egy blokk tridiagonális mátrix lesz:

$$P = \begin{pmatrix} B_0 & C_1 & 0 & 0 & 0 & 0 \\ A_0 & B_1 & C_2 & 0 & 0 & 0 \\ 0 & A_1 & B_2 & C_3 & 0 & 0 \\ 0 & 0 & A_2 & B_3 & C_4 & 0 \\ \vdots & \vdots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & A_{N-1} & B_N \end{pmatrix}.$$

Az alapegyenletek pedig a következő alakban adhatóak meg:

$$\dot{x}_k = A_{k-1}x_{k-1} + B_kx_k + C_{k+1}x_{k+1}, \quad k = 0, 1, \dots, N,$$

ahol $A_{-1} = 0$ és $C_{N+1} = 0$, és az $A_k \in \mathbb{R}^{d_{k+1} \times d_k}$ mátrix elemei az $m_1 \rightarrow m_2$ átmenetért felelős ráták, míg $C_k \in \mathbb{R}^{d_{k-1} \times d_k}$ elemei a $m_2 \rightarrow m_1$ átmenet rátái. A B_k egy diagonális mátrix főátlójában a $-(e_{k+1}A_k + e_{k-1}C_k)$ vektorral, ahol $e_k = (1, \dots, 1)$ egy d_k dimenziós sorvektor.

Az általános modell egy speciális gyakran vizsgált esetével fogunk foglalkozni a továbbiakban, hogy részleteiben is megértsük a folyamat viselkedését.

2.4. A klasszikus SIS modell

Az egyik legegyszerűbb betegségterjedési modell az SIS típusú dinamika, ahol S jelöli az egészséges, míg I a beteg állapotot, az állapotter 2^N elemű és $I \rightarrow S$ illetve $S \rightarrow I$ típusú átmenet lehetséges. A G szimmetrikus szomszédsági mátrixban tároljuk, hogy mely csúcsok között vezet él, azaz $G = \{0, 1\}^{N \times N}$. Az $S = \{S, I\}^N$ állapotteret ebben az esetben is érdemes a fentebb említett $S = \cup_{k=0}^N S^k$ felbontásban vizsgálni. Tehát az S állapotteret felosztjuk $N+1$ részhalmazzra a következő módon:

- $S^0 := \{S_1^0\}$, ahol az S^0 részhalmaz tartalmazza a gráf 0 darab beteget tartalmazó állapotait. Látható, hogy ilyen állapotból csak egy van: $S_1^0 = (S \dots S)$.
- $S^k := \{S_1^k, \dots, S_{d_k}^k\}$, S^k tartalmazza a gráf azon állapotait, ahol pontosan k darab beteg van. Ilyen állapotokból $d_k = \binom{N}{k}$ darab létezik, $k = 0, \dots, N$.
- $S^N := \{S_1^N\}$, azaz az S^N halmazban legyenek a gráf N darab beteget tartalmazó állapotai, ami azt jelenti, hogy az egy darab ilyen állapot: $S_1^N = (I \dots I)$.

Most ismertetünk egy gyakran használt és jól áttekinthető $I \rightarrow S$ és $S \rightarrow I$ átmeneti rátákat megadó függvény családot, de megjegyezzük, hogy ettől eltérő modelleket is szép számmal lehet találni az irodalomban.

Legyen a meggyógyulás rátája

$$f_{IS}(n, G, S_i^k) := \gamma \quad (2.3)$$

bármely $n \in \{1, \dots, N\}$ esetén, tetszőleges S_i^k állapot mellett, feltéve hogy $S_i^k(n) = I$, tehát a n -edik csúcs beteg és $\gamma \geq 0$ egy adott paraméter, amivel a meggyógyulás erősségét lehet szabályozni. Ez azt jelenti, hogy a meggyógyulás rátája független a csúcs pozíciójától a gráfban, illetve a gráf állapota se befolyásolja. Ezzel szemben a megbetegedés rátája $\tau \geq 0$ paraméter mellett legyen

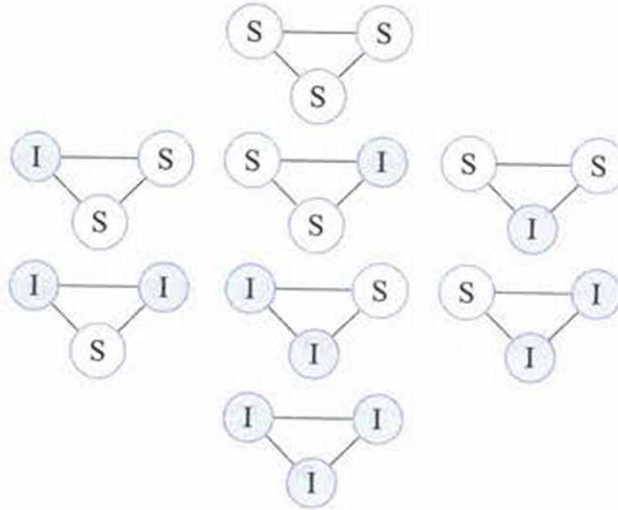
$$f_{SI}(n, G, S_i^k) := \tau \cdot \sum_{\{p \in \{1, \dots, N\} | S_i^k(p) = I\}} g_{pn}, \quad (2.4)$$

feltéve hogy $S_i^k(n) = S$, vagyis az n -edik csúcs egészséges. Itt látható, hogy a megbetegedés rátája függ a gráfstruktúrától, a folyamat jelenlegi állapotától és az egészséges csúcs helyzetétől is. Ez a típusú átmeneti ráta röviden úgy fogalmazható, hogy feltéve, hogy az n -edik csúcs egészséges és l beteg szomszédal rendelkezik, a megbetegedési rátája: $l\tau$, hiszen $\sum_{\{p=1, \dots, N | S_i^k(p) = I\}} g_{pn}$ pontosan az n -edik csúcs beteg szomszédainak a számát adja.

A (2.2) alapegyenlethez tartozó P együtthatómátrix természetesen ezen dinamika mellett is blokktridiagonális alakú lesz. A mátrix elemeiről pontos tájékoztatást a [26] cikkben találhatunk.

2.4.1. A klasszikus modell bemutatása 3 csúcsú teljes gráf esetén

Az előbbi betegségterjedési modellt egy három csúcsú teljes gráfon illusztráljuk. Tehát legyen $N = 3$, így az állapottér: $S = \{S, I\}^3$, és az állapottér felbontása a betegek száma alapján a



2.1. ábra. A 3 csúcsú teljes gráf lehetséges állapotai.

szokásos részhalmazokra a következő (ld. 2.1. ábra):

- $S^0 := \{S_1^0\} = \{(SSS)\},$
- $S^1 := \{S_1^1, S_2^1, S_3^1\} = \{(ISS), (SIS), (SSI)\},$
- $S^2 := \{S_1^2, S_2^2, S_3^2\} = \{(IIS), (ISI), (SII)\},$
- $S^3 := \{S_1^3\} = \{(III)\}.$

A Markov-folyamathoz kapcsolódó alapegyenlet-rendszer a következő formát ölti:

$$\begin{aligned}
 \dot{x}_{SSS}(t) &= \gamma(x_{SSI}(t) + x_{SIS}(t) + x_{ISS}(t)), \\
 \dot{x}_{ISS}(t) &= \gamma(x_{IIS}(t) + x_{ISI}(t)) - (2\tau + \gamma)x_{ISS}(t), \\
 \dot{x}_{SIS}(t) &= \gamma(x_{IIS}(t) + x_{SII}(t)) - (2\tau + \gamma)x_{SIS}(t), \\
 \dot{x}_{SSI}(t) &= \gamma(x_{ISI}(t) + x_{SII}(t)) - (2\tau + \gamma)x_{SSI}(t), \\
 \dot{x}_{IIS}(t) &= \gamma x_{III}(t) + \tau(x_{ISS}(t) + x_{SIS}(t)) - 2(\tau + \gamma)x_{IIS}(t), \\
 \dot{x}_{ISI}(t) &= \gamma x_{III}(t) + \tau(x_{ISS}(t) + x_{SSI}(t)) - 2(\tau + \gamma)x_{ISI}(t), \\
 \dot{x}_{SII}(t) &= \gamma x_{III}(t) + \tau(x_{SIS}(t) + x_{SSI}(t)) - 2(\tau + \gamma)x_{SII}(t), \\
 \dot{x}_{III}(t) &= 2\tau(x_{IIS}(t) + x_{ISI}(t) + x_{SII}(t)) - 3\gamma x_{III}(t),
 \end{aligned} \tag{2.5}$$

ahol az egyszerűség kedvéért a továbbiakban az x_{ABC} jelölést alkalmazzuk az $x_{(ABC)}$ jelölés helyett, $A, B, C \in \{S, I\}$. A (2.5) egyenletrendszerben például a második egyenlet esetén

az látható, hogy az (*ISS*) állapotba az (*IIS*) és az (*ISI*) állapotból lehet bekerülni γ rátájú gyógyulás által, illetve kimenni az (*ISS*) állapotból τ rátájú fertőződéssel lehet az (*IIS*) illetve az (*ISI*) állapotokba, vagy gyógyulással az (*SSS*) állapotba, tehát $2\tau + \gamma$ azon ráták összege, amik az (*ISS*) állapotból kivezetnek. Az egyenletrendszer felírható a (2.2) $\dot{x}(t) = P x(t)$ mátrixos alakban is, ahol

$$P = \begin{pmatrix} 0 & \gamma & \gamma & \gamma & 0 & 0 & 0 & 0 \\ 0 & -(2\tau + \gamma) & 0 & 0 & \gamma & \gamma & 0 & 0 \\ 0 & 0 & -(2\tau + \gamma) & 0 & \gamma & 0 & \gamma & 0 \\ 0 & 0 & 0 & -(2\tau + \gamma) & 0 & \gamma & \gamma & 0 \\ 0 & \tau & \tau & 0 & -2(\tau + \gamma) & 0 & 0 & \gamma \\ 0 & \tau & 0 & \tau & 0 & -2(\tau + \gamma) & 0 & \gamma \\ 0 & 0 & \tau & \tau & 0 & 0 & -2(\tau + \gamma) & \gamma \\ 0 & 0 & 0 & 0 & 2\tau & 2\tau & 2\tau & 3\gamma \end{pmatrix},$$

és

$$x(t) = (x_{SSS}(t), x_{ISS}(t), x_{SIS}(t), x_{SSI}(t), x_{IIS}(t), x_{ISI}(t), x_{SII}(t), x_{III}(t))^T$$

az állapotok valószínűségeit tartalmazó oszlopvektor. Jelen esetben 2^3 darab alapegyenletet kapunk. Mivel 2 lehetséges állapota van a csúcsoknak, így a P mátrix blokk tridiagonális, tehát

$$P = \begin{pmatrix} B_0 & C_1 & 0 & 0 \\ A_0 & B_1 & C_2 & 0 \\ 0 & A_1 & B_2 & C_3 \\ 0 & 0 & A_2 & B_3 \end{pmatrix} \quad (2.6)$$

alakú, ahol

$$B_0 = \begin{pmatrix} 0 \end{pmatrix}, \quad C_1 = \begin{pmatrix} \gamma & \gamma & \gamma \end{pmatrix},$$

$$A_0 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \quad B_1 = \begin{pmatrix} -(2\tau + \gamma) & 0 & 0 \\ 0 & -(2\tau + \gamma) & 0 \\ 0 & 0 & -(2\tau + \gamma) \end{pmatrix}, \quad C_2 = \begin{pmatrix} \gamma & \gamma & 0 \\ \gamma & 0 & \gamma \\ 0 & \gamma & \gamma \end{pmatrix},$$

$$A_1 = \begin{pmatrix} \tau & \tau & 0 \\ \tau & 0 & \tau \\ 0 & \tau & \tau \end{pmatrix}, \quad B_2 = \begin{pmatrix} -2(\tau + \gamma) & 0 & 0 \\ 0 & -2(\tau + \gamma) & 0 \\ 0 & 0 & -2(\tau + \gamma) \end{pmatrix}, \quad C_3 = \begin{pmatrix} \gamma \\ \gamma \\ \gamma \end{pmatrix},$$

$$A_2 = \begin{pmatrix} 2\tau & 2\tau & 2\tau \end{pmatrix}, \quad B_3 = \begin{pmatrix} 3\gamma \end{pmatrix}.$$

Nagyon fontos észrevétel, hogy *SIS* dinamika esetén a (2.3) és (2.4) átmenetvalószínűségek mellett a $\dot{x}(t) = P x(t)$ alapegyenletek jobb oldalából ha kiemelünk γ -t, akkor a P együttható mátrix elemei már csak τ/γ lineáris függvényei lesznek. Ez jól látszik az előbbi 3 csúcsú teljes gráf P mátrixa esetén, szemléltetésül az együttható mátrix egy része, a B_2 mátrix alakja a kiemelés után:

$$B_2 = \gamma \begin{pmatrix} -2(\frac{\tau}{\gamma} + 1) & 0 & 0 \\ 0 & -2(\frac{\tau}{\gamma} + 1) & 0 \\ 0 & 0 & -2(\frac{\tau}{\gamma} + 1) \end{pmatrix}.$$

Ezek alapján a (2.2) alapegyenletek $\dot{x}(t) = \gamma \hat{P} x(t)$ alakúra módosulnak, ahol $P = \gamma \hat{P}$. Ez azt jelenti hogy csak a τ/γ aránytól függenek a betegségterjedés lezajlásának a különböző sajátosságai, például a kvázi-stacionárius állapot értéke, a γ paraméter pedig az időskálát állítja be, azaz a folyamat lejátszódásának a gyorsaságát. Ezt a megfigyelést a későbbiekben még használni fogjuk, illetve tovább pontosítjuk.

2.4.2. Alacsony dimenziós modellek

A (2.2) alapegyenlet-rendszer kis N esetén numerikusan megoldható, azonban ahogy növeljük a gráf méretét az alapegyenletek száma a csúcsszám függvényében exponenciálisan nő, így ezen egyenletek felírása és a megoldások numerikus kiszámítása reménytelen, ezért a megoldásfüggvények hiányában a folyamatról se lesz megfelelő mennyiségű információnk. Ez indokolja azt a törekvést, hogy a nagy méretű alapegyenlet-rendszer helyett lecsökkentett méretű közelítő egyenleteket használjunk, melyek megbecsülik a folyamat néhány fontos jellemzőjét, például az átlagos betegszámot. Számos ilyen modell létezik, de a legegyszerűbb az átlagolással kapott egyenlet (mean-field):

$$[\dot{I}](t) = \tau[SI](t) - \gamma[I](t), \quad (2.7)$$

ahol $[I](t)$ jelöli, hogy a t időpontban a gráf átlagosan hány beteg csúcsot tartalmaz, illetve $[SI](t)$ az SI élek számának várható értékét adja meg a t időpontban. Formálisan az olvasható le az egyenletből, hogy a betegszám növekedése az SI élek átlagos számától függ τ rátával, míg a betegszámmal arányosan csökken γ rátával. A (2.7) egyenlet a (2.2) alapegyenletekből pontosan levezethető [26], azonban az egyenlet nincs lezárva, azaz mivel az $[SI]$ és az $[I]$ változók is ismeretlenek, így nem oldható meg a differenciálegyenlet. Az egyenlet egy lehetséges lezárása:

$$[SI] \approx \frac{n}{N-1} [I](N - [I]), \quad (2.8)$$

ahol az SI élek átlagos számát $[I]$ függvényében becsüljük és n az átlagos fokszám. A közelítés mögötti kombinatorikai érvelés szerint egy beteg csúcs egy kimenő éle $\frac{N-[I]}{N-1}$ valószínűséggel kapcsolódik egészséges csúcshoz, melyek átlagos száma a gráfban $[S] = N - [I]$, így egy átlagosan n éllel rendelkező beteg csúcsból várhatóan $n \frac{N-[I]}{N-1}$ darab SI él indul ki, ezt szorozzuk meg a betegek átlagos számával, hogy kapjuk egy becslést $[SI]$ -re. A lezárás után így az $\dot{[I]} = \tau \frac{n}{N-1} [I](N - [I]) - \gamma [I]$ differenciálegyenletet kapjuk, azonban az SI élek átlagos számára vonatkozó becslés nem annyira pontos a különböző gráftípusok esetén. Ennek az az oka, hogy a (2.8) becslés egyenletesen elszórt beteg csúcsokat feltételez a gráfon, pedig a betegségterjedés következtében a fertőzött csúcsok nem egyenletesen oszlanak el a gráfban, hanem nagyobb valószínűséggel kapcsolódnak egymáshoz, mint az egészséges csúcsokhoz. Ezért a lezárt egyenlet megoldása se feltétlenül közelíti jól a betegek átlagos számát a különböző gráfokon, lásd 2.2. ábra. A modell pontosítására egy lehetőség, hogy megtartjuk az $[SI]$ változót és felírunk arra is egy differenciálegyenletet, illetve felírunk egy-egy egyenletet az $[SS]$ és az $[II]$ változókra is, azaz az SS élek és az II élek átlagos számára is. Ez a párok szintjén felírt rendszer (pairwise):

$$\begin{aligned} \dot{[I]}(t) &= \tau [SI](t) - \gamma [I](t), \\ \dot{[S]}(t) &= \gamma [I](t) - \tau [SI](t), \\ \dot{[SI]}(t) &= \gamma ([II](t) - [SI](t)) + \tau ([SSI](t) - [ISI](t) - [SI](t)), \\ \dot{[II]}(t) &= -2\gamma [II](t) + 2\tau ([ISI](t) + [SI](t)), \\ \dot{[SS]}(t) &= 2\gamma [SI](t) + 2\tau [SSI](t). \end{aligned} \quad (2.9)$$

Megjegyezzük, hogy az első és a második egyenlet egymással ekvivalens, felhasználva az $[I] + [S] = N$ összefüggést, ugyanígy az utolsó három egyenlet is összefügg, hiszen $[SS] + 2[SI] +$

$[II] = nN$ fennáll a gráfban található összes lehetséges élek összegére. Az első két egyenlet a (2.7) összefüggés kibővítve az egészségesek átlagos számára felírt differenciálegyenlettel. Ezek az átlagolt egyenletek vannak kiegészítve a párokra fennálló differenciálegyenletekkel. Ezekben az egyenletekben megjelennek a hármasok, mint új változók, ezekre kell egy lezáró képletet adni. A szokásos közelítő összefüggés:

$$[ABC] \approx \frac{n-1}{n} \frac{[AB][BC]}{[B]},$$

ahol A, B, C jelöli a csúcsok típusait, tehát $A, B, C \in \{S, I\}$ és n az átlagos fokszám [18]. A becslés azon a heurisztikus érvelésen alapszik, miszerint először vizsgáljuk meg, hogy átlagosan hány A típusú szomszédja van egy B csúcshoz, ezekből válasszunk ki egyet, majd a B csúcs maradék $n-1$ éléből nézzük meg, hogy átlagosan hány BC típusú. Az AB és BC típusú élek átlagos számának szorzata adja egy B csúcsához kapcsolódó $[ABC]$ hármasok átlagos számát. Pontosan végigkövetve a gondolatot, egy B csúcsból átlagosan $n[B]$ él indul ki és $[AB]$ az AB típusú élek átlagos száma, így egy B csúcsból kiinduló AB élek aránya: $\frac{[AB]}{n[B]}$, illetve a B csúcsból kiinduló AB élek átlagos száma: $n \frac{[AB]}{n[B]} = \frac{[AB]}{[B]}$, ebből kiválasztunk egyet. A maradék $n-1$ él esetén ugyanilyen érveléssel a BC típusú élek aránya: $\frac{[BC]}{n[B]}$, átlagos száma: $(n-1) \frac{[BC]}{n[B]}$, ezekből is kiválasztunk egyet. Így egy konkrét B csúcson keresztül haladó ABC hármasok átlagos száma:

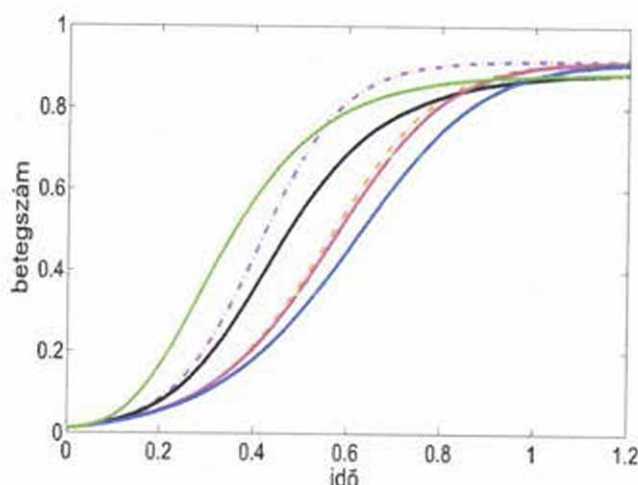
$$\frac{[AB]}{[B]} (n-1) \frac{[BC]}{n[B]},$$

tehát $[B]$ átlagos darabszámú B típusú csúcs esetén a gráfban lévő ABC hármasok átlagos száma: $\frac{n-1}{n} \frac{[AB][BC]}{[B]}$. Ezek alapján a (2.9) egyenletben szereplő hármasokra a

$$[SSI] \simeq \frac{n-1}{n} \frac{[SS][SI]}{[S]} \quad \text{és} \quad [ISI] \simeq \frac{n-1}{n} \frac{[SI]^2}{[S]} \quad (2.10)$$

becsléseket kapjuk. Természetesen más becslésekkel is lezárható az egyenlet [17], a lényeg hogy a hármasokat a párok és a betegszám várható értékének segítségével becsüljük meg. Az ilyen formulákat momentum lezárási képleteknek nevezik.

A (2.9) rendszer a (2.7) egyenlettel szemben, már nem csak az átlagos betegszámra ad egy differenciálegyenletet, hanem a különböző típusú párok számáról is információt szolgáltat. A (2.9) rendszer megoldása pontosabb a (2.7) egyenlet megoldásánál, de még így is meglehetősen pontatlan sok gráftípus esetén. Ennek az az oka, hogy a gráfstruktúra nincs beleszámítva az egyenletekbe, csak az átlagos fokszám jelenik meg a képletekben, de a fokszámeloszlás egyéb



2.2. ábra. A betegek arányának várható értéke a teljes populációban a szimulációból Barabási-Albert gráf (zöld), bimodális gráf (fekete), reguláris véletlen gráf (piros) és klaszterezett gráf (kék) esetén (0.4-es klaszterezettségi együtthatóval) illetve a (2.7) egyenlet (lila szaggatott görbe) és a (2.9) egyenletek (barna szaggatott görbe), $N = 1000$, $n = 6$, $\tau = 2$, $\gamma = 1$. 250 szimuláció átlagából, 10 kezdeti beteggel.

jellemzői nem. Tehát azt lehet mondani, hogy homogén foksámeloszlású gráf esetén ezek az alacsony dimenziós közelítő egyenletek jó becslést adhatnak a folyamatot jellemző változókra, de heterogén foksámeloszlású vagy speciális struktúrájú gráfok esetén bonyolultabb modellekre van szükség. Ezt a 2.2. ábrán szemléltetjük. Látható, hogy a (2.7) egyenlet a vizsgált gráfokon nem írja jól le a folyamat működését, azonban a (2.9) rendszer rendszer a 6 átlagos foksámmal rendelkező reguláris véletlen gráfon kifejezetten pontos eredményt ad.

A (2.2) alapegyenletek számának lecsökkentésére egy másik lehetséges módszer az egyenletek összevonása a gráf automorfizmusai alapján, ami speciális struktúrájú gráfok esetén hatékony eljárást adhat. Az S állapottér elemei között egy ekvivalencia relációt határozhatunk meg olyan módon, hogy ekvivalensnek tekintjük azon állapotokat, melyek automorfizmussal egymásba vihetők. Ennek a technikának az a lényege, hogy összeadjuk a folyamat azon állapottaihoz tartozó egyenleteket, melyek egy ekvivalencia osztályba tartoznak. Így egy csökkentett méretű egzakt $\dot{y} = By$ egyenletrendszert kapunk, melynek legyen B az együttható mátrixa és y az összevont állapotosztályok valószínűségeit tartalmazó K dimenziós oszlopvektor, ahol K az ekvivalencia osztályok száma. Az eljárás során definiáljuk az un. $T \in \{0, 1\}^{K \times 2^N}$ összevonás mátrixot, tehát legyen $T_{i,j} = 1$ pontosan akkor, ha a Markov-folyamat j -edik állapota

az i -edik összevont állapotosztályban található, egyébként $T_{i,j} = 0$. Ekkor az eredeti (2.2) alapegyenlet-rendszert leíró P és az összevont rendszerhez tartozó B együttható mátrixok között fennáll a $TP = BT$ összefüggés. Az $\dot{y} = By$ rendszer megoldásából kiszámítható a (2.2) alapegyenlet-rendszer megoldása az $x = T^T y$ formula segítségével. Sajnos ez a technika csak nagyon speciális gráfok esetén vezet csak hatékony eljáráshoz, ugyanis a legtöbb esetben a gráf összes automorfizmusa nehezen adható meg, illetve az összevonás nem csökkenti jelentősen a rendszer méretét. Erről a módszerről bővebben a [26] cikkben olvashatunk.

2.5. A numerikus szimuláció algoritmus

A szimuláció algoritmusát $M = 2$ estén fogjuk vázolni, azaz abban az esetben, ha bármely csúcs két lehetséges állapotban lehet, azonban ez az algoritmus könnyedén általánosítható tetszőleges véges méretű állapothalmazra is.

$M = 2$ esetén az állapottér az $\{m_1, m_2\}^N$ halmaz. Az eddigieknek megfelelően $X(t) \in \{m_1, m_2\}^N$ jelöli azt az N dimenziós vektort, ami megadja, hogy a folyamat a t időpontban melyik állapotban van. Ha az $X(t)$ vektor n -edik koordinátája $X_n(t) = m_1$, azaz az n -edik csúcs m_1 állapotú, akkor $1 - e^{-f_{12}\delta t}$ a valószínűsége annak, hogy δt idő alatt az m_2 állapotba kerül, ahol $f_{12} = f_{12}(n, G, X(t))$. Ugyanígy ha az n -edik csúcs m_2 állapotú, akkor $1 - e^{-f_{21}\delta t}$ a valószínűsége annak, hogy δt idő alatt az m_1 állapotba kerül, ahol $f_{21} = f_{21}(n, G, X(t))$.

Ezek alapján a numerikus szimuláció a következőképpen zajlik: megadjuk előre a bemeneti paramétereket, azaz a δt időlépést, az időlépések számát, a G adjacencia mátrixot, egy $X(0)$ kezdeti feltételt és a futtatások ismétlésszámát. Tegyük fel, hogy az algoritmussal eljutottunk az $X(t)$ állapotig és nézzük meg, hogyan hajtható végre a következő időlépés. Ehhez az időlépéshez generálunk egy $v \in [0, 1]^N$ véletlen vektort, aminek minden koordinátája egy véletlen szám a $[0, 1]$ intervallumból. A v vektor n -edik eleme legyen v_n , ennek a segítségével döntjük el, hogy az n -edik csúcs állapotában bekövetkezik-e változás. Kihasználva, hogy kellően kicsi δt választással az $1 - e^{-x\delta t} \approx x\delta t$ becslés jó közelítést ad, az n -edik csúcs m_1 állapotból m_2 állapotba jut δt idő alatt, ha

$$v_n < f_{12}\delta t,$$

amennyiben $X_n(t) = m_1$; és ugyanígy m_2 állapotból m_1 állapotba kerül δt idő alatt, ha

$$v_n < f_{21}\delta t,$$

feltéve, hogy $X_n(t) = m_2$.

Tehát egy időlépés során minden csúcs esetén kiszámítjuk az állapotváltozás rátáját az $X(t)$ állapotvektor és a G szomszédsági mátrix alapján, majd a v véletlen vektor felhasználásával eldöntjük az összes csúcs esetében, hogy δt idő alatt történik-e állapotváltozás. Az $X(t)$ állapotvektor koordinátáit ezek alapján megváltoztatjuk, így előállítva az új $X(t + \delta t)$ állapotvektort. Az ismert algoritmust futtatva megkapjuk a Markov-folyamat egy trajektóriáját. Tipikusan a folyamat valamely jellemző adatára vagyunk kíváncsiak az idő függvényében, mint például az átlagos betegszám, vagy az SI élek várható értéke. Ezeket az értékeket az egyes futtatások során kapott trajektóriákból számítjuk ki. Amennyiben kellően sokszor lefuttattuk a szimulációt - létrehozva az ismétlésszám mennyiségű realizációt -, ezek átlagolása után a (2.2) alapegyenletek által leírt folyamat egy kifejezetten jó közelítéséhez jutunk, csak arra figyelni kell, hogy a δt időlépést megfelelően kicsire válasszuk és egy idő lépés alatt legfeljebb egy csúcs állapota változzon meg, ugyanis ellenkező esetben a szimuláció nem pontosan azt a folyamatot modellezné, amit a 2.3. szakaszban megadtunk.

Az algoritmuson egy kicsit változtatni kell, ha véletlen gráfokkal megadott hálózatokon dolgozunk, ilyen típusú modellekre a 3. fejezetben láthatunk példát. Véletlen gráfok esetén egy \mathcal{G} halmazzal adjuk meg azon típusú gráfok osztályát, melyeken a folyamatot értelmezhetjük, illetve a \mathcal{K} halmazban tároljuk a lehetséges kezdeti feltétel vektorokat. A szimuláció kezdetekor ezen halmazokon kívül megadjuk a δt időlépést, az időlépések számát és a futtatások ismétlésszámát is. Minden egyes konkrét futtatáskor kiválasztunk véletlenszerűen egy gráfot a \mathcal{G} halmazból egyenletes eloszlás szerint, illetve egy kezdeti feltételt a \mathcal{K} halmazból, végül ezek alapján kiszámítjuk a folyamat egy trajektóriáját az előbb ismert algoritmus segítségével. A folyamat realizációja alapján a kívánt mennyiségeket könnyen kiszámolhatjuk, majd az eljárást az ismétlésszám eléréséig ismétljük, végül a kapott eredményeket átlagoljuk. Megjegyezzük, hogy a \mathcal{G} halmazt úgy határozzuk meg, hogy a kijelölt típusú gráf összes lehetséges megvalósulása pontosan egyszer szerepeljen \mathcal{G} -ben, illetve a \mathcal{K} halmaz is pontosan egyszer tartalmazza az összes lehetséges kezdeti feltételt. A szimulálás során egy gráfnak a \mathcal{G} halmazból való kiválasztását úgy is kivitelezhetjük, hogy egy megfelelő gráf generáló algoritmussal előállítjuk a véletlen gráftípus egy konkrét realizációját, majd azon futtatjuk az algoritmust.

A numerikus szimuláció egy hasznos eszköz a Markov-folyamat viselkedésének vizsgálatához. Egyrészt a futtatás során előállítjuk a folyamat különböző realizációit, amik pontosabb

képet adnak a folyamat működéséről és ez motiválhatja a kutatás irányát, illetve ötletet adhat a felmerülő problémák megoldására. Másrészt megnézhetjük, hogy a kínált modellek megoldása mennyire jól közelíti a szimulációs értéket, ezzel ellenőrizve a módszer pontosságát. A numerikus szimuláció hátránya, hogy analitikus vizsgálatra alkalmatlan, ezért általános törekvés analitikus közelítő módszerek kidolgozása.

3. fejezet

Járványterjedés modellezése véletlen gráfokkal megadott hálózatokon

3.1. Motiváció

Tekintsünk egy N csúcsú egyszerű gráfot SIS típusú dinamika mellett. Most a 2.4. szakaszban ismertetett átmenetvalószínűségekkal dolgozunk: egy beteg csúcs γ rátával gyógyul meg, míg egy egészséges csúcs l beteg szomszéd esetén τl rátával kapja el a fertőzést. A hálózat és a dinamika meghatározása után egy 2^N méretű állapottérrel rendelkező Markov-folyamatot kapunk, melyhez kapcsolódóan felírható a (2.2) alapegyenlet-rendszer. Az értekezés egyik célkitűzése, az alapegyenletek számának olyan módú redukálása, ami minél pontosabban megőrzi a folyamatot jellemző mennyiségeket, fő tekintettel a betegek számának várható értékére.

Ebben a fejezetben abból a jól ismert tényből indulunk ki, hogy egy N csúcsú teljes gráf esetén a 2^N dimenziós alapegyenletek száma $N + 1$ -re csökkenthető úgy, hogy a k beteget tartalmazó állapotokat - függetlenül azok elhelyezkedésétől - egy állapotnak, precízen a k -adik állapotnak tekintjük a csökkentett $\{0, 1, \dots, N\}$ állapottéren [26]. Az így kapott redukált egyenletrendszer pontosan modellezi a folyamatot és

$$\dot{x}_k(t) = a_{k-1}x_{k-1}(t) - (a_k + c_k)x_k(t) + c_{k+1}x_{k+1}(t), \quad k = 0, 1, \dots, N \quad (3.1)$$

alakú, ahol $x_k(t)$ jelöli annak a valószínűségét, hogy k beteg csúcs található a gráfban a t időpontban, illetve az együtthatók

$$a_k = \tau k(N - k), \quad c_k = \gamma k, \quad k = 0, \dots, N \quad \text{és} \quad a_{-1} = c_{N+1} = 0. \quad (3.2)$$

Heurisztikusan ez azt jelenti, hogy az egészséges csúcsok γ rátával gyógyulhatnak meg, így k beteg esetén γk a gyógyuláshoz tartozó együttható. A beteg csúcsok τ rátával fertőződhetnek meg a gráfban található SI éleken keresztül, melyek száma $k(N - k)$, ez magyarázza a fertőződésért felelős $a_k = \tau k(N - k)$ együtthatót. A redukált differenciálegyenlet-rendszeren is jól látható, hogy a k beteges állapotba kétféleképpen lehet eljutni: a $k - 1$ beteget tartalmazó állapotból fertőződéssel, míg a $k + 1$ beteges állapotból gyógyulás által. A teljes gráf szimmetriája teszi lehetővé az egyenletek számának az extrém lecsökkentését, ugyanis a teljes gráf esetén minden k beteget tartalmazó állapot automorfizmussal egymásbavihető, vagyis az \mathcal{S}^k állapot-halmaz összes eleme egy ekvivalencia osztályba tartozik, ezért nem érdekes a beteg csúcsok pontos elhelyezkedése, csak azok száma befolyásolja a folyamat alakulását. Megjegyezzük, hogy teljes gráf esetén a (3.1) egyenletek a (2.2) alapegyenletekből precízen levezethetők, hiszen összeadva a k beteges állapotokhoz tartozó $d_k = \binom{N}{k}$ darab egyenletet a (3.1) rendszert kapjuk. Erről a technikáról a következő részben részletesen írunk.

Felmerül a kérdés, hogy ugyanígy le lehet-e csökkenteni az alapegyenletek számát $N + 1$ -re tetszőleges struktúrájú gráf esetén is, esetleg akár csak úgy, hogy a csökkentés a pontosság rovására megy és csupán egy közelítő differenciálegyenlet-rendszert kapunk. Két fő kérdéskört lehet megfogalmazni ezen ötlet kapcsán. Az egyik, hogy lehetséges-e megszámlálni, vagy megbecsülni az SI élek átlagos számát a betegszám függvényében, ugyanis ezt a becslést fogjuk felhasználni az a_k együtthatóknál, a másik pedig, hogy a becsült együtthatókkal megoldott redukált egyenletrendszer jó egyezést mutat-e az alapegyenlet-rendszer megoldásával vagy a szimulációval. Adott betegszám mellett az SI élek számának kikalkulálása egyáltalán nem triviális, ugyanis figyelembe kell venni, hogy a fertőzött csúcsok elhelyezkedését a gráfon a betegségterjedés folyamata határozza meg, amit a gráf struktúrája erősen befolyásol. Még ha ez a lépés sikeres is és előállítjuk az a_k együtthatókat, nincs arra garancia, hogy a redukált alapegyenletek megoldása jól közelíti a pontos modellt.

A fejezet célja, hogy információt szerezzünk a járványterjedés lefolyásáról az idő függvényében, így az átlagos betegszámot szeretnénk megbecsülni különböző struktúrájú gráfok esetén. Csak olyan paraméter beállításokat használunk, amik mellett kis betegszámból indítva a betegség végigterjed a gráfon, azaz olyan eseteket nem vizsgálunk, amikor a betegség megszűnik az idő előrehaladtával. A megfelelő paraméter beállításokról információt az 5. fejezetben találhatunk. A figyelem középpontjába klasszikus gráftípusokat helyeztünk, mint az

Erdős-Rényi gráf, a reguláris véletlen gráf, a bimodális véletlen gráf illetve a negatív binomiális foksámeloszlású gráf, a Barabási-Albert gráf és a klaszterezett gráf, amit egy reguláris véletlen gráfból nyerünk a klaszterezetségi szintet növeleve egy ún. Big V rewiring technikával [17]. Csak a felsorolt gráftípusokon alkalmazzuk az eljárást, azonban tetszőleges foksámeloszlású, konfigurációs modellel előállítható gráf esetén is alkalmazható módszert kínálunk. A fő kihívás az a_k fertőződési ráta meghatározása, ahol a_k/τ az SI élek átlagos számát kell, hogy megadja. Az eljárás során heurisztikus és kombinatorikai összefüggéseket is felhasználtunk, hogy meghatározzuk az említett értékeket.

A fejezet felépítése a következő. Először bevezetjük a redukált egyenletrendszert és megmutatjuk, hogy azt a szimulációból kimért átmeneti rátákkal megoldva az átlagos betegszámra kapott görbe tökéletes egyezést mutat a szimulációs eredményekkel, így a (2.2) alapegyenletrendszerral is. A 3.3. szakaszban bemutatjuk a kombinatorikai alapokon nyugvó eljárásunkat az II élek átlagos számának meghatározására, abban az esetben, ha a gyógyulás lehetőségét kizárjuk, azaz SI dinamikát tekintünk. Ez a módszer megjósolja az II élek átlagos számát tetszőleges konfigurációs gráf esetén. Erre alapozva, a 3.4. szakaszban meghatározzuk az SI élek várható értékét, és kiterjesztjük az eljárást kis τ esetre is felhasználva a (2.9) rendszert. Reguláris véletlen gráf esetén egy explicit formulát is felírunk az SI élek átlagos számára a betegszám függvényében. A módszerünk lépéseiről az 3.5. szakaszban adunk egy áttekintést.

3.2. A modell formalizálása

Összefoglalva a fenti gondolatokat, a 2^N méretű S állapottér helyett az egyszerűbb $\{0, 1, \dots, N\}$ állapottéren fogunk dolgozni. Ez az új Markov-folyamat sokkal átláthatóbb, ugyanis a k -adik állapotból, ami a k beteges állapotot jelöli, csak a $k+1$ -edik állapotba mehetünk fertőződéssel, illetve a $k-1$ -edik állapotba gyógyulás által. A kapcsolódó redukált egyenletrendszer:

$$\dot{x}_k(t) = a_{k-1}x_{k-1}(t) - (a_k + c_k)x_k(t) + c_{k+1}x_{k+1}(t), \quad k = 0, 1, \dots, N, \quad (3.3)$$

ahol megint csak $x_k(t)$ jelöli annak a valószínűségét, hogy k beteg csúcs található a gráfban a t időpontban. Emellett egyrészt $c_k = \gamma k$ a gyógyulásért felelős együttható, ugyanis egy beteg csúcs a szomszédai állapotától és a gráf struktúrájától teljesen függetlenül gyógyul meg, csak a γ ráta befolyásolja a meggyógyulás gyorsaságát. Másrészt $a_k = \tau e_{SI}(k)$, ahol $e_{SI}(k)$ jelöli

az SI élek átlagos számát k beteg esetén, hiszen az SI éleken keresztül fertőződhetnek meg az egészséges csúcsok.

A továbbiakban megvizsgáljuk, hogy a (2.2) alapegyenletekből, hogyan lehet levezetni a (3.3) redukált egyenletrendszert, avagy a 2^N méretű S állapottérből hogyan juthatunk el a $\{0, 1, \dots, N\}$ állapottérhez. Ehhez elevenítsük fel a bevezetőben ismertetett jelöléseket. A 2^N dimenziós alapegyenlet-rendszer

$$\dot{\mathbf{x}}(t) = \mathbf{P}\mathbf{x}(t)$$

alakú, és a $d_k = \binom{N}{k}$ dimenziós \mathbf{x}_k oszlopvektorban tároljuk rendre a k beteges állapotok valószínűségeit

$$\mathbf{x}_k := (x_{S_1^k}, x_{S_2^k}, \dots, x_{S_{d_k}^k})^T, \quad k = 0, \dots, N.$$

Legyen $S^k = \{S_1^k, \dots, S_{d_k}^k\}$ a k beteges állapotokat tartalmazó halmaz. Mint már említettük a 2.3. szakaszban, az alapegyenletek

$$\dot{\mathbf{x}}_k = A_{k-1}\mathbf{x}_{k-1} + B_k\mathbf{x}_k + C_{k+1}\mathbf{x}_{k+1}, \quad k = 0, 1, \dots, N,$$

alakba is írhatóak, ahol az A_k mátrixok elemei az S^k állapothalmazból az S^{k+1} állapotaiba menő állapotváltozások átmeneti rátái, azaz a fertőződésért felelnek, míg a C_k mátrixok a gyógyulás szabályait rögzítik és $A_{-1} = 0$, $C_{N+1} = 0$. A B_k egy olyan diagonális mátrix, melynek főátlója a $-(e_{k+1}A_k + e_{k-1}C_k)$ vektor, ahol $e_k = (1, \dots, 1)$ egy d_k dimenziós sorvektor. Az alapegyenletek számát úgy fogjuk csökkenteni, hogy az S^k -ba tartozó állapotokat összeolvasztjuk.

Teljes gráf esetén az alapegyenletek redukcióját a k beteges állapotokhoz tartozó egyenletek összeadásával lehet elérni, így ugyanezt fogjuk tenni általános esetben is. Tekintsük az alapegyenleteket és adott k -ra az S^k -beli állapotok differenciálegyenleteit adjuk össze, tehát vegyük a $\dot{\mathbf{x}}_k = A_{k-1}\mathbf{x}_{k-1} + B_k\mathbf{x}_k + C_{k+1}\mathbf{x}_{k+1}$ egyenletek összegét. Legyen

$$\mathbf{x}_k := e_k \mathbf{x}_k = x_{S_1^k} + x_{S_2^k} + \dots + x_{S_{d_k}^k}.$$

Ekkor az új változókra fennáll az

$$\dot{\mathbf{x}}_k = e_k A_{k-1} \mathbf{x}_{k-1} + e_k B_k \mathbf{x}_k + e_k C_{k+1} \mathbf{x}_{k+1}$$

differenciálegyenlet. Felhasználva B_k ismert szerkezetét az egyenlet a következő alakba is írható:

$$\dot{\mathbf{x}}_k = e_k A_{k-1} \mathbf{x}_{k-1} - (e_{k+1} A_k + e_{k-1} C_k) \mathbf{x}_k + e_k C_{k+1} \mathbf{x}_{k+1}.$$

Mivel a C_{k+1} mátrix minden oszlopában pontosan $k+1$ darab γ van és a többi elem nulla [26], így $e_k C_{k+1} = (k+1)\gamma e_{k+1}$, tehát

$$e_k C_{k+1} \mathbf{x}_{k+1} = (k+1)\gamma e_{k+1} \mathbf{x}_{k+1} = (k+1)\gamma x_{k+1}.$$

Ami azt jelenti, hogy $c_k = k\gamma$, $k = 0, 1, \dots, N$.

Másrészt tudjuk, hogy a A_{k-1} mátrixok oszlopaiban elhelyezkedő τ elemek száma megegyezik az oszlophoz tartozó állapotban található SI élek számával [26], azaz $e_k A_{k-1} = \tau \cdot N_{SI}(\mathcal{S}^{k-1})$, ahol

$$N_{SI}(\mathcal{S}^{k-1}) = (N_{SI}(\mathcal{S}_1^{k-1}), N_{SI}(\mathcal{S}_2^{k-1}), \dots, N_{SI}(\mathcal{S}_{d_k}^{k-1}))$$

egy d_k dimenziós sorvektor, és $N_{SI}(\mathcal{S}_j^{k-1})$ az \mathcal{S}_j^{k-1} állapotban lévő SI élek száma.

Mindezek alapján az új változókra felírt differenciálegyenlet-rendszer a következő formát ölti:

$$\dot{x}_k = \tau \cdot N_{SI}(\mathcal{S}^{k-1}) \mathbf{x}_{k-1} + e_k B_k \mathbf{x}_k + (k+1)\gamma x_{k+1}.$$

Hogy megkapjuk a (3.3) redukált rendszert, meg kell határoznunk azt az a_k együtthatót melyre pontosan, vagy legalább közelítőleg fennáll a

$$\tau \cdot N_{SI}(\mathcal{S}^k) \mathbf{x}_k = a_k x_k$$

összefüggés. Először nézzük meg, hogy a legegyszerűbb esetben, vagyis a teljes gráf esetén hogyan határozhatjuk meg az a_k értékeket. Az \mathcal{S}_j^k állapotban, ami k beteget tartalmaz, az SI élek száma $k(N-k)$, $j = 1, 2, \dots, d_k$, azaz - függetlenül attól, hogy az \mathcal{S}^k halmaz melyik elemét vettük, - az SI élek száma állandó. Vagyis

$$N_{SI}(\mathcal{S}^k) = k(N-k)e_k. \quad (3.4)$$

Amiből azt kapjuk, hogy

$$N_{SI}(\mathcal{S}^k) \mathbf{x}_k = k(N-k)e_k \mathbf{x}_k = k(N-k)x_k,$$

tehát teljes gráf esetén $a_k = \tau k(N-k)$, ahogy vártuk a (3.2) képlet alapján.

Tetszőleges struktúrájú gráf esetén az SI élek száma az \mathcal{S}_j^k állapotokban különféle lehet különböző j értékek mellett, tehát a k beteget tartalmazó egyes állapotokban. Így a (3.4) összefüggés helyett csak egy közelítő képletet van esély felírni:

$$N_{SI}(\mathcal{S}^k) \approx e_{SI}(k)e_k, \quad (3.5)$$

valamilyen megfelelően választott mesterséges $e_{SI}(k)$ értékkel, amit a továbbiakban az SI élek átlagos számának nevezünk k beteg esetén. Ezek után feltételezve, hogy meghatározhatóak az $e_{SI}(k)$ értékek valamilyen módon $k = 0, 1, \dots, N$ esetén, az a_k együtthatókra teljesülni fog, hogy

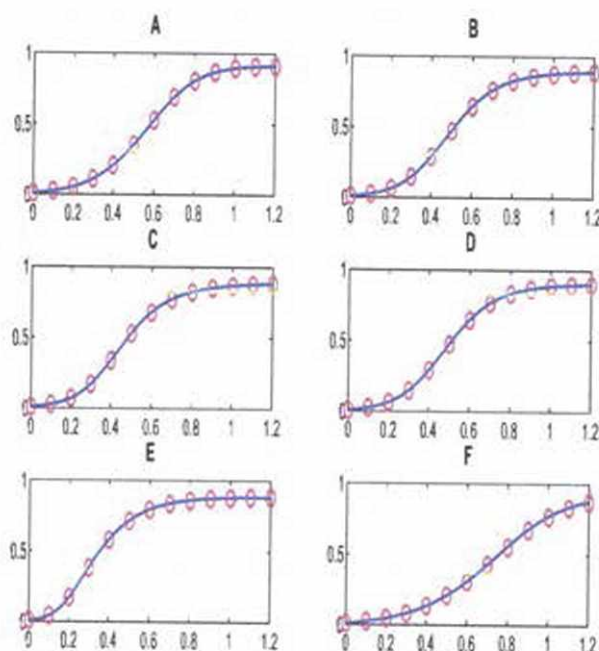
$$a_k = \tau e_{SI}(k). \quad (3.6)$$

A következő részben megvizsgáljuk, hogy érdemes-e a redukált differenciálegyenlet-rendszer felírásával foglalkozni, majd a továbbiakban egy jól használható eljárást adunk az $e_{SI}(k)$ értékek, azaz az SI élek átlagos számának meghatározására a betegszám függvényében.

3.2.1. A modell megvalósíthatósága

Azt hihetnénk, hogy a 2^N méretű állapottér lecsökkentése a $\{0, 1, \dots, N\}$ állapothalmazra, túldrasztikus méretvesztés, így nem lehet leírni a folyamatot megfelelően. Azonban meglepően nagy pontosságot lehet elérni az új állapottéren is, amit a következőképpen támasztunk alá. A 3.1. ábrákon összevetettük a szimulációból kimért betegszám-arányt a redukált egyenletekből kapott betegszám-aránnyal, ahol az a_k együtthatókat a szimulációból vettük, azaz $a_k = \tau e_{SI}(k)$, ahol $e_{SI}(k)$ most kivételesen a szimuláció során átlagosan megjelenő SI élek számát jelöli k beteg esetén. A két grafikon tökéletes egyezése az összes vizsgált gráfípus esetén megfigyelhető, ami azt mutatja, hogy ha meg tudnánk adni analitikusan is az SI élek átlagos számát rögzített betegszám mellett, akkor egy olyan elméleti módszerhez jutnánk, amelyben a (3.3) redukált rendszer jól modellezné a folyamatot.

Mindezek alapján a fő kihívás az a_k ráták meghatározása, pontosabban az SI élek átlagos számának a kiszámítása a betegszám függvényében. Ezen mennyiségek szimulációból történő kimérése során eltekintünk attól, hogy az SI élek száma igazából egy valószínűségi változó valamilyen eloszlással, és csak a várható értékét számoljuk ki. Ugyanígy az elméleti becslés meghatározásakor is csak a várható értékre leszünk kíváncsiak. Adott betegszám esetén az SI élek pontos száma sokféle értéket felvehet attól függően, hogy a fertőzött csúcsok milyen elrendezésben helyezkednek el a \mathcal{S}^k halmaz különböző állapotaiban, így ez a változatosság adja a feladatnak a nehézségét. A továbbiakban heurisztikus és kombinatorikai érveléssel adunk egy képletet az SI élek átlagos számára.



3.1. ábra. A betegek arányának várható értéke a teljes populációban a szimulációból (\circ) és a szimulációból kimért e_{SI} értékekkel megoldott (3.3) redukált egyenletekből (kék görbe) reguláris (A), Erdős-Rényi (B), bimodális (C), negatív binomiális (D), Barabási-Albert (E) és klaszterezett (F) gráfok esetén. (A klaszterezett gráf 0.4-es klaszterezettségi együtthatóval szerepel.) $N = 1000$, $n = 6$, $\tau = 2$, $\gamma = 1$. 250 szimuláció átlagából, 10 kezdeti beteggel.

3.2.2. A fertőződésért felelős átmeneti ráták meghatározása

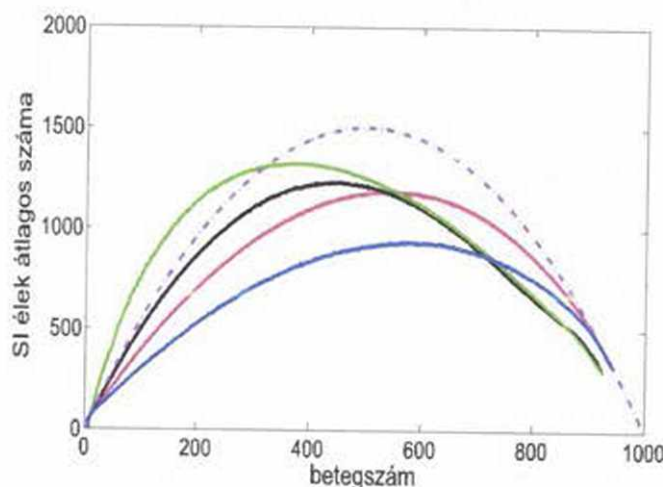
Mint már említettük, az egyik leggyakrabban használt elméleti becslés az a_k együtthatókra

$$a_k = \tau n k \frac{N - k}{N - 1}, \quad (3.7)$$

ahol n jelöli az átlagos fokszámot. Ez a (2.8) becslés k beteg esetén adódó alakja és az a problémája, hogy a fertőzött csúcsok egyenletes elosztását feltételezi a gráfon, holott a a beteg csúcsok egymáshoz kapcsolódva jelennek meg a járványterjedés következtében.

A 3.2. ábrán a reguláris véletlen gráf, a bimodális gráf, a Barabási-Albert gráf és a klaszterezett gráf szimulációból kimért e_{SI} értékeit vetjük össze ezzel az elméleti közelítéssel. Jól látható, hogy a (3.7) becslés erősen felülbecsüli az SI élek átlagos számát, illetve hogy a gráfstruktúra erősen befolyásolja a valós e_{SI} értékeket. Megfigyelhető például a Barabási-Albert gráf esetén, hogy a fokszámeloszlás erősödő heterogenitása az e_{SI} görbe maximumát balra és egyre feljebb tolja, mivel a magas fokszámú csúcsokhoz, azaz a középpontokhoz kis betegszám mellett is hamar elér a megbetegedés, és ezek a fertőzött csúcsok sok egészségeshez kapcsolódnak, ami megnöveli az SI élek számát. Másrészt egy magas klaszterezettségi együtthatójú gráfnál a görbe maximuma jobbra és lefelé tolódik, hiszen ilyenkor egy beteg csúcs valószínűbb, hogy egy másik beteg csúcshoz kapcsolódik a gyakoribb háromszögek miatt, ami csökkenti az SI élek számát.

A (3.7) becslés gyengesége arra sarkall, hogy egy összetettebb közelítést adjunk az SI élek átlagos számára, ami jobban visszatükrözi a gráf struktúráját és sokkal pontosabb eredményekhez vezet. Ugyanis a gráfstruktúra sincs beleszámítva a (3.7) közelítő képletbe, csupán egyetlen - a gráf típusát jellemző - paramétert, az átlagos fokszámot használja fel, ami minimális információ a valós fokszámeloszlás figyelembevételéhez képest. Ez azt jelenti, hogy ez a becslés tetszőleges típusú, n átlagos fokszámmal rendelkező gráfra ugyanazt az eredményt adja. Az a_k együtthatókba bele kell építeni a gráfstruktúra több jellemzőjét is valamilyen formában. A kidolgozott elmélet alapját az II élek átlagos számának meghatározása adja, melyet a következő fejezetben mutatunk be, majd az II élekre kapott képletből következtetünk az SI élek átlagos számára.



3.2. ábra. Az $e_{SI}(k)$ értékek a szimulációból Barabási-Albert gráf (zöld), bimodális gráf (fekete), reguláris véletlen gráf (piros) és klaszterezett gráf (kék) esetén illetve a (3.7) becslés (szaggatott görbe), $N = 1000$, $n = 6$, $\tau = 2$, $\gamma = 1$. 250 szimuláció átlagából, 10 kezdeti beteggel.

3.3. Elméleti becslés az II élek várható értékére

Tehát az SI élek átlagos száma helyett, most az II élek átlagos számára írunk fel először egy formulát, ugyanis a beteg részgráf kialakulását fogjuk nyomon követni, ami minden egyes új beteg hozzávétele esetén újabb II éleket hoz létre. Az átláthatóság kedvéért szeretnénk kizárni a visszagyógyulás lehetőségét a fertőzött részgráfban, ezért ebben a fejezetben SI dinamikát feltételezünk, amit majd a szimulációval való összevetés során úgy fogunk elérni, hogy a τ/γ arányt megfelelően nagyra választjuk. Ahogy a bevezetőben már említettük, csupán ez a hányados már meghatározza a betegségterjedés lefolyását, így csak ez az arány befolyásolja az újonnan megjelenő II élek átlagos számát és ebből kifolyólag az e_{SI} értékeket is, amit majd a későbbiekben ábrával is szemléltetünk.

A betegségterjedést követve építjük majd fel a gráfot a konfigurációs modell szabályai alapján, mivel a klaszterezett gráftól eltekintve az összes gráftípus - megadva a kívánt fokszámeloszlást - konfigurációs modellel generáltnak tekinthető. A betegségterjedési folyamat mentén konstruált gráf a szimulációkból nyert tapasztalat szerint megfelel a konfigurációs modellel előállított gráfnak.

Megadunk egy rekurzív formulát $e_{II}(k)$ értékeire, ahol $e_{II}(k)$ jelöli az II élek várható ér-

tékét k beteg esetén. Ez a rekurzió jól működik minden olyan struktúrájú gráf esetén, amit a konfigurációs modellel lehet előállítani. Az első részben a rekurziós képlet felírását a legegyszerűbb esetben, tehát egy reguláris véletlen gráfon mutatjuk be. Ezt követően a rekurziós formulát általánosítjuk tetszőleges foksámeloszlású gráfra.

3.3.1. A reguláris véletlen gráf esete

Vegyünk egy reguláris véletlen gráfot, aminek minden csúcsa n fokú. Tegyük fel, hogy a gráfban k darab beteg csúcs található és ismerjük $e_{II}(k)$ értékét, azaz az II élek átlagos számát. Ezekből a feltételezésekből kiindulva azt fogjuk vizsgálni, hogy egy új fertőződés hatására mennyivel nő az II élek átlagos száma, és ez alapján meg fogunk adni egy explicit formulát az $e_{II}(k+1) - e_{II}(k)$ különbségre.

1. Állítás. Tekintsünk egy N csúcsú reguláris véletlen gráfot n átlagos foksámmal, illetve $e_{II}(k)$ jelölje az II élek várható értékének egy közelítését k beteg esetén. Ekkor $e_{II}(1) = 0$ és

$$e_{II}(k+1) - e_{II}(k) = 2 \frac{n(N-k) + (n-1)(nk - e_{II}(k)) - n}{n(N-k) - 1}, \quad k = 1, 2, \dots, N-1. \quad (3.8)$$

BIZONYÍTÁS. Tegyük fel, hogy k beteg csúcs van a gráfban, a beteg részgráfot már összekapcsoltuk és az egészséges csúcsokat most fogjuk hozzákapcsolni a beteg részgráf szabad fél-éleihez. Vezessünk be néhány új jelölést. A k beteg csúcsból $m := nk - e_{II}(k)$ darab fél-él indul ki, ami az egészséges csúcsokhoz kapcsolódhat, ugyanis átlagosan $e_{II}(k)$ darab, a beteg csúcsokból kiinduló fél-él köti össze a beteg részgráf csúcsait. Az egészséges csúcsokból kimenő fél-élek száma $M := n(N-k)$, hiszen $N-k$ az egészséges csúcs van. Ezeket a fél-éleket szeretnénk összekapcsolni a konfigurációs modell szabályai alapján. Ezt úgy képzeljük el, hogy a beteg csúcshalmaz szabad fél-éleit hozzácsatlakoztatjuk az egészséges csúcsokhoz, még hozzá úgy, hogy az M egészséges fél-élből kiválasztjuk azt az m darabot, amit a beteg csúcsokhoz kötünk be. Ezt $\binom{M}{m}$ -féleképpen tehetjük meg. Itt fontos megjegyeznünk, hogy azt nem kell számon tartani, hogy a beteg csúcsok között hogyan futnak az élek, csak annyit kell tudnunk hogy $e_{II}(k)$ darab beteg fél-él már foglalt. Az eljárás során kizárólag a beteg csúcsok és az egészséges csúcsok közötti SI élek létrehozásának folyamatát kell vizsgálni.

A továbbiakhoz lényeges meghatároznunk, hogy mi a valószínűsége annak, hogy egy tetszőleges egészséges csúcs j éllel csatlakozik a beteg csúcshalmazhoz, jelöljük ezt a valószínűséget

p_j -vel. Látható, hogy p_j hipergeometrikus eloszlást követ, hiszen egy egészséges csúcs n fél-élből j darabot kell kiválasztani, amit a beteg csúcsokhoz kötünk, majd a maradék egészséges csúcsok $M - n$ fél-élből $m - j$ darabot, hogy szintén a beteg csúcshalmazba fussanak be. Tehát

$$p_j = \frac{\binom{n}{j} \binom{M-n}{m-j}}{\binom{M}{m}}, \quad j = 0, 1, \dots, n.$$

Mivel p_j jelöli a j fertőzött szomszédal rendelkező egészséges csúcsok arányát, azért az ilyen tulajdonsággal rendelkező csúcsok átlagos száma $(N - k)p_j$. Felhasználva, hogy a hipergeometrikus eloszlás várható értéke $\frac{nm}{M}$, az új élek behúzásával kapott SI élek átlagos száma

$$(N - k) \sum_{j=0}^n j p_j = (N - k) \frac{nm}{M} = (N - k) \frac{nm}{n(N - k)} = m,$$

ami természetesen megegyezik a beteg csúcsok szabad fél-éleinek a számával. A kalkuláció során az új II élek átlagos számát kell meghatároznunk. Ezt az értéket az alapján lehet kiszámolni, hogy megvizsgáljuk, hogy melyik SI élen fog keresztülhaladni a betegség, és melyik egészséges csúcs fog ezáltal megfertőződni, majd ezek után azt kell megnézni, hogy ennek a már megfertőzött csúcsnak hány beteg szomszédja volt a megbetegedése pillanatában.

Tehát először tekintsük az SI éleket és mindegyikhez rendeljünk hozzá egy számot, úgy hogy pontosan akkor legyen ez a szám j , ha az adott SI él egészséges csúcsának j beteg szomszédja van. Ez azt jelenti, hogy egy egészséges csúcs minden kimenő SI élére a j címkét helyezzük, ha az adott egészséges csúcsnak j beteg szomszédja van. Így a j jelöléssel ellátott SI élek átlagos száma $j(N - j)p_j$, ugyanis átlagosan $(N - k)p_j$ darab j beteg szomszédal rendelkező egészséges csúcs van és mindegyikből j darab SI él megy ki. Könnyen látható, hogy annak a valószínűsége, hogy j darab új II él keletkezik az új fertőződés következtében, megegyezik annak az eséllyel, hogy a felcímkézett SI élek közül pont egy olyanon keresztül történik a megbetegedés, amin a j jelzés szerepel. Ez a valószínűség

$$q_j = \frac{j(N - j)p_j}{m}, \quad j = 1, 2, \dots, n.$$

Ezek alapján az új fertőződés következtében keletkező új II élek számának várható értéke: $2 \sum j q_j$, ugyanis II típusú élek esetén egy élet mindkét irányból meg kell számolnunk. Vagyis

$$e_{II}(k + 1) - e_{II}(k) = 2 \sum_{j=1}^n j q_j = 2 \sum_{j=1}^n \frac{j^2(N - j)p_j}{m} = 2 \frac{N - k}{m} \sum_{j=1}^n j^2 p_j.$$

A képletben szereplő összeg a hipergeometrikus eloszlás második momentuma, ami kiszámítható a hipergeometrikus eloszlás szórásnégyzetének (D^2) és várható értékének (E) segítségével:

$$\sum_{j=1}^n j^2 p_j = D^2 + E^2,$$

ahol $E = \frac{nm}{M}$ és $D^2 = \frac{nm}{M} \frac{M-n}{M} \frac{M-m}{M-1}$. Tehát

$$\sum_{j=1}^n j^2 p_j = \frac{nm}{M} \frac{M-n}{M} \frac{M-m}{M-1} + \left(\frac{nm}{M}\right)^2,$$

ami alapján

$$e_{II}(k+1) - e_{II}(k) = 2 \frac{N-k}{m} \frac{nm}{M} \left(\frac{M-n}{M} \frac{M-m}{M-1} + \frac{nm}{M} \right) = 2 \frac{M-m-n+mn}{M-1}.$$

Behelyettesítve az $m = nk - e_{II}(k)$ és $M = n(N-k)$ jelöléseket a fenti képletbe, megkapjuk a rekurziós összefüggést:

$$e_{II}(k+1) - e_{II}(k) = 2 \frac{n(N-k) + (n-1)(nk - e_{II}(k)) - n}{n(N-k) - 1}, \quad k = 1, 2, \dots, N-1.$$

A rekurzió az $e_{II}(1) = 0$ értékkel indul, ugyanis egy beteg esetén az II élek száma nulla. \square

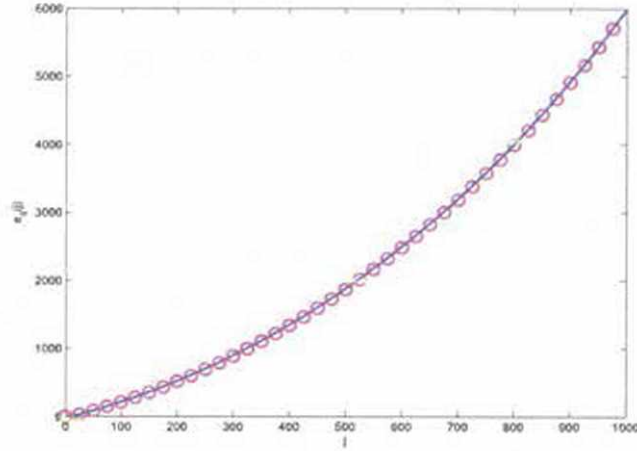
A 3.3. ábrán a rekurzióból számított $e_{II}(k)$ értékeket illetve a szimulációból kimért $e_{II}(k)$ értékeket hasonlítjuk össze. Az ábrán jól látható a kitűnő egyezés az elméleti és a szimulációs értékek között. Ez az eredmény motiválja, hogy kiterjesszük a rekurziós eljárást heterogén fokszámoszlású gráfok esetére is. A következő fejezet ezzel foglalkozik.

3.3.2. Tetszőleges fokszámoszlású gráf esete

Tekintsünk egy N csúcsú véletlen gráfot és legyen N_l az n_l fokú csúcsok száma, azaz a gráf fokszámoszlása: $p(n_l) = N_l/N$, $l = 1, 2, \dots, L$, ahol L a különböző fokok száma, emellett $N_1 + N_2 + \dots + N_L = N$. Heterogén fokszámoszlás esetén számon kell tartani, hogy adott betegszám esetén mennyi a különböző fokú beteg csúcsok átlagos száma. Legyen k a betegek száma a gráfban, míg $I_l(k)$ az n_l fokú beteg csúcsok számának várható értéke. Nyilván fennáll a $\sum_{l=1}^L I_l(k) = k$ egyenlőség.

Először megadunk egy rekurziós összefüggést $I_l(k)$ -re. Tegyük fel, hogy adott k esetén ismerjük a $I_l(k)$ értékeket, $l = 1, 2, \dots, L$. Jelöljük $P_l(k)$ -vel annak a valószínűségét, hogy k beteg esetén - a következő fertőzött csúcs foka n_l , ekkor az n_l fokú beteg csúcsok számának várható értéke $k+1$ beteg esetén:

$$I_l(k+1) = I_l(k) + P_l(k).$$



3.3. ábra. Az $e_{II}(k)$ értékek a szimulációból (◊) és a rekurzióból (kék görbe) reguláris véletlen gráf esetén, $N = 1000$, $n = 6$, $\tau = 10$, $\gamma = 1$. 250 szimuláció átlagából, 10 kezdeti beteggel.

A $P_l(k)$ valószínűség kiszámítására később visszatérünk, most megvizsgáljuk, hogy átlagosan hány új II él keletkezik egy új csúcs megbetegedésekor, azaz megadunk egy formulát az $e_{II}(k+1) - e_{II}(k)$ különbségre.

Az, hogy átlagosan hány új II él keletkezik egy új csúcs megbetegedésekor, attól függ, hogy a $k+1$ -edik új beteg csúcsnak mi a foka. Ezért először a feltételes várható értékét számítjuk ki az $e_{II}(k+1) - e_{II}(k)$ kifejezésnek, abban az esetben, ha az új fertőzött foka n_l , majd a teljes valószínűség tételét felhasználva kapjuk meg az újonnan keletkező II élek átlagos számát.

A levezetés az előző fejezet lépéseivel analóg módon történik. Feltesszük, hogy k beteg van a gráfban és ezekből

$$m = \sum_{l=1}^L n_l I_l(k) - e_{II}(k) \quad (3.9)$$

fél-él áll ki, amit az egészséges csúcsokhoz csatlakoztathatunk. A $\sum_{l=1}^L (N_l - I_l(k))$ darab egészséges csúcsnak

$$M = \sum_{l=1}^L n_l (N_l - I_l(k)) \quad (3.10)$$

fél-éle van összesen. Most azt az esetet vizsgáljuk meg, amikor a következő megbetegedő csúcs foka n_l . Legyen p_j^l annak a valószínűsége, hogy egy tetszőleges n_l fokú egészséges csúcsnak j beteg szomszédja van. Ekkor p_j^l hipergeometrikus eloszlást követ:

$$p_j^l = \frac{\binom{n_l}{j} \binom{M-n_l}{m-j}}{\binom{M}{m}}, \quad j = 0, 1, \dots, n_l.$$

Azt kaptuk, hogy $(N_l - I_l(k))p_j^l$ azon n_l fokú egészséges csúcsoknak az átlagos száma, melyeknek j beteg szomszédja van. Ezek alapján az n_l fokú egészséges csúcsokból kiinduló SI élek várható értéke

$$(N_l - I_l(k)) \sum_{j=0}^n j p_j^l = (N_l - I_l(k)) \frac{n_l m}{M},$$

ahol felhasználtuk, hogy a hipergeometrikus eloszlás várható értéke $\frac{n_l m}{M}$.

Mivel feltettük, hogy a következő megbetegedő csúcs n_l fokú, ezért csak az n_l fokú egészséges csúcsokból kiinduló SI éleket címkézzük most fel egy számmal. Hasonlóan az előző esethez, az SI élen a hozzárendelt érték legyen az él egészséges csúcsához kapcsolódó beteg szomszédok száma. Így a j címkével ellátott SI él átlagos száma $j(N_l - I_l(k))p_j^l$, hiszen $(N_l - I_l(k))p_j^l$ az n_l fokú, j beteg szomszédal rendelkező egészséges csúcsoknak a száma, és minden ilyen csúcsból j él megy ki fertőzött csúcsokba. Ezen SI élek egyikén keresztül fog megfertőződni a $k + 1$ -edik beteg csúcs, és ha ezen az SI élen a j szám szerepel, akkor j új II él fog keletkezni a gráfban átlagosan, ennek a valószínűségét jelöljük q_j^l -val. Tehát feltéve, hogy a $k + 1$ -edik beteg csúcs foka n_l , a keresett valószínűség a következő formulával kapható meg:

$$q_j^l = \frac{j(N_l - I_l(k))p_j^l}{(N_l - I_l(k)) \frac{n_l m}{M}} = \frac{M}{n_l m} j p_j^l, \quad j = 1, 2, \dots, n_l,$$

ugyanis a kedvező SI élek számát kell elosztanunk az összes n_l fokú egészséges csúcsból kiinduló SI élek számával.

Ezek alapján a keletkező II élek várható értékét egy n_l fokú új csúcs megbetegedésekor a következő módon kaphatjuk meg:

$$E_l(k) = \sum_{j=1}^{n_l} j q_j^l = \frac{M}{n_l m} \sum_{j=1}^{n_l} j^2 p_j^l = \frac{M - n_l}{M} \frac{M - m}{M - 1} + \frac{n_l m}{M},$$

tehát itt egy feltételes várható értéket írtunk fel, és felhasználtuk az előző fejezetből, hogy

$$\sum_{j=1}^{n_l} j^2 p_j^l = \frac{n_l m}{M} \frac{M - n_l}{M} \frac{M - m}{M - 1} + \left(\frac{n_l m}{M} \right)^2.$$

Térjünk most vissza $P_l(k)$ kiszámításához, ami annak a valószínűségét adja meg, hogy k beteg csúcs esetén a következő fertőzött csúcs n_l fokú. Ez a valószínűség egyszerűen kiszámítható, hiszen $P_l(k)$ az n_l fokú egészséges csúcsokból kiinduló SI élek aránya az összes SI élhez képest:

$$P_l(k) = \frac{(N_l - I_l(k)) \frac{n_l m}{M}}{\sum_{l=1}^L (N_l - I_l(k)) \frac{n_l m}{M}} = \frac{(N_l - I_l(k)) n_l}{\sum_{l=1}^L (N_l - I_l(k)) n_l}.$$

Végül a rekurziós formula $e_{II}(k)$ -re megkapható, ha felhasználjuk a teljes várható érték tételét:

$$e_{II}(k+1) - e_{II}(k) = 2 \sum_{l=1}^L P_l(k) E_l(k), \quad k = 1, \dots, N-1.$$

Ebben az esetben is az $e_{II}(1) = 0$ értékből indítható a rekurzió.

Összefoglalva az előbbi levezetést, kimondható az alábbi állítás.

2. Állítás. *Tekintsünk egy tetszőleges foksámeloszlású gráfot. A fenti jelölések mellett az II élek várható értékére a következő rekurzió adódik:*

$$e_{II}(k+1) - e_{II}(k) = 2 \sum_{l=1}^L P_l(k) E_l(k), \quad k = 1, \dots, N-1, \quad e_{II}(1) = 0, \quad (3.11)$$

ahol

$$P_l(k) = \frac{(N_l - I_l(k))n_l}{\sum_{l=1}^L (N_l - I_l(k))n_l}, \quad k = 1, \dots, N-1, \quad l = 1, \dots, L,$$

és

$$E_l(k) = \frac{M - n_l}{M} \frac{M - m}{M - 1} + \frac{n_l m}{M}, \quad k = 1, \dots, N-1, \quad l = 1, \dots, L.$$

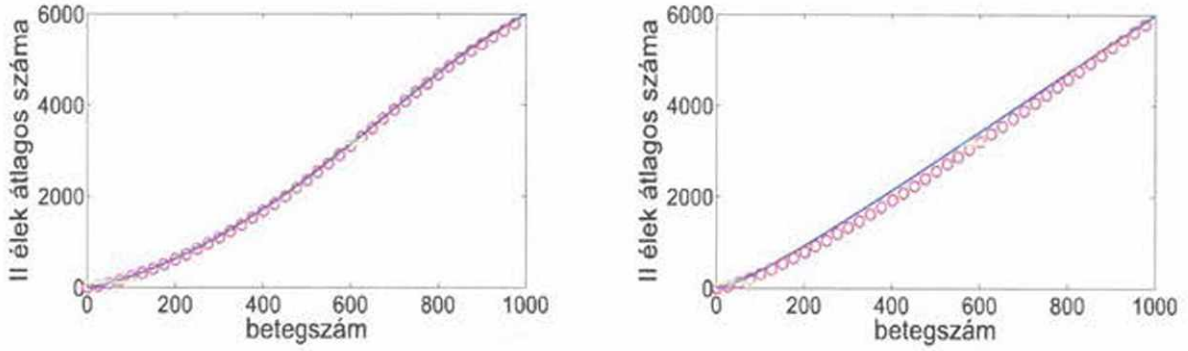
Illetve a különböző fokú betegek számára felírt rekurzió:

$$I_l(k+1) = I_l(k) + P_l(k), \quad k = 1, \dots, N-1, \quad l = 1, \dots, L.$$

Az elméletet először egy 1000 csúcsú bimodális gráfon próbáltuk ki. A csúcsok felének a foka 3, a másik felének pedig 9, azaz $L = 2$, $n_1 = 3$, $n_2 = 9$, $N_1 = 500 = N_2$. A (3.11) rekurzió segítségével kiszámoltuk $e_{II}(k)$ értékeit, amiket a szimulációból kimért $e_{II}(k)$ értékekkel hasonlítottunk össze. Ugyanezt az eljárást Barabási-Albert gráfon is megismérteltük. Az eredmények a 3.4. ábrán láthatóak. A szimulációs eredmények azt mutatják, hogy az elméleti becslés hatékony eljárást ad heterogén foksámeloszlású gráfon is.

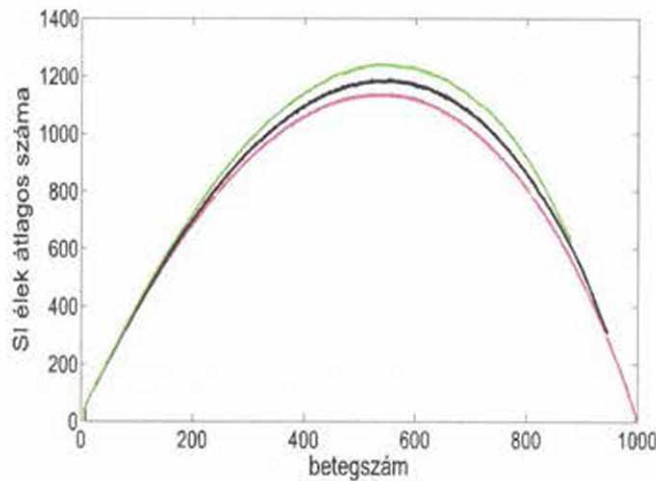
3.4. Elméleti becslés az SI élek várható értékére

Ebben a szakaszban megadunk egy közelítő képletet az $e_{SI}(k)$ értékekre, azaz az SI élek várható értékére k beteg esetén, felhasználva az II élek átlagos számára kapott becslést. A szimulációból arra lehet következtetni, hogy a $(k, e_{SI}(k))$ ($k = 0, 1, \dots, N$) pontok egy parabolászerű görbét követnek, amit a 3.5. ábrán is láthatunk. Az e_{SI} görbe alakja a gráfstruktúrától függ, míg a görbe maximum értéke a τ/γ aránytól, ezért a továbbiakban γ értékét lerögzítjük 1-re, és



3.4. ábra. Az $e_{II}(k)$ értékek a szimulációból (○) és a rekurzióból (kék görbe) bimodális véletlen gráf esetén (baloldali ábra) és Barabási-Albert gráf esetén (jobboldali ábra), $N = 1000$, $n = 6$, $\tau = 10$, $\gamma = 1$. 250 szimuláció átlagából, 10 kezdeti beteggel.

τ függvényében vizsgáljuk az e_{SI} görbe maximum értékének változását: ahogy τ értéke növekszik, a görbe maximum értéke csökken. A szimuláció alapján egyre nagyobb τ értékekre az e_{SI} görbe egy határgörbéhez tart, melyet $e_{SI}^{\infty}(k)$ -vel jelölünk, utalva a $\tau \rightarrow \infty$ határátmenetre. A véges határgörbe megjelenése azzal magyarázható, hogy ha a τ értéke megfelelően nagy, vagy másképp fogalmazva a τ/γ arány elér egy értéket, akkor a folyamat során elhanyagolhatóan kicsi lesz a gyógyulás esélye, azaz lényegében ilyenkor SIS helyett SI típusú dinamika jelenik meg a gráfon. Az e_{SI} görbe maximum értékének a τ -tól való függését a 3.5. ábrán láthatjuk.



3.5. ábra. A $(k, e_{SI}(k))$ görbék a szimulációból reguláris véletlen gráf esetén, $N = 1000$, $\gamma = 1$. 250 szimuláció átlagából, 10 kezdeti beteggel, különböző τ értékekre: $\tau = 1$ (zöld), $\tau = 2$ (fekete), $\tau = 10$ (piros).

Felhasználva az eddigi megfigyeléseinket, két lépésben építjük fel az elméleti modellt e_{SI} meghatározására tetszőleges τ mellett. Először megadunk egy képletet $e_{SI}^\infty(k)$ -ra, itt feltesszük, hogy τ annyira nagy a γ -hoz képest, hogy lényegében gyógyulás nem következhet be. Majd megvizsgáljuk, hogy hogyan befolyásolja τ megválasztása az e_{SI} görbe maximum értékét.

Az első lépéshez annyit kell észrevennünk, hogy a gráfban megtalálható SI és II élek összege megegyezik a fertőzött csúcsokból kiinduló fél-élek számával, vagyis

$$e_{SI}^\infty(k) = \sum_{l=1}^L n_l I_l(k) - e_{II}(k), \quad k = 0, \dots, N,$$

ahol $e_{II}(k)$ a rekurzióból kiszámolt érték. Például a reguláris véletlen gráf esetén, ahol minden csúcs foka n , ez a képlet az alábbi egyszerű formát ölti:

$$e_{SI}^\infty(k) = nk - e_{II}(k).$$

Tehát az előző fejezetben levezetett rekurzió az $e_{II}(k)$ értékekre egyből megadja $e_{SI}^\infty(k)$ -t. Itt megjegyezzük, hogy a ∞ szimbólumot azért nem használjuk $e_{II}(k)$ esetén, mert az $e_{II}(k)$ -ra felírt rekurzió levezetése alatt végig nagy τ -t feltételeztünk, avagy SI dinamikát, ezzel szemben $e_{SI}(k)$ értékeit kis τ esetén is szeretnénk kiszámítani a későbbiekben, így itt már meg kell különböztetni a nagy és a kis τ esetét.

Habár $e_{II}(k)$ -ra és $e_{SI}^\infty(k)$ -ra megadtunk egy rekurziós képletet, explicit formulát sajnos nem tudunk felírni. Azonban nagy N esetén egy közelítő explicit képletet kaphatunk reguláris véletlen gráf esetén, ha a rekurziós összefüggés segítségével felírunk egy differenciálegyenletet az SI élek átlagos számára a betegek arányának függvényében. Ezt a gondolatot fejtjük ki részletesebben a következő alszakaszban.

3.4.1. Explicit formula az $e_{SI}^\infty(k)$ értékekre reguláris véletlen gráf esetén

Az alszakaszban a rekurziós összefüggésből kiindulva megadunk egy explicit formulát az SI élek átlagos számára reguláris véletlen gráf esetén.

3. Állítás. Tekintsünk egy N csúcsú reguláris véletlen gráfot, és legyen minden csúcs foka n . Felhasználva a (3.8) formulát, e_{SI}^∞ előállítható

$$e_{SI}^\infty(k) = nN \left(1 - \frac{k}{N}\right) - n(N-2) \left(\frac{nN(1 - \frac{k}{N}) - 1}{n(N-1) - 1} \right)^{2-2/n}. \quad (3.12)$$

alakban.

BIZONYÍTÁS. Alkalmazva az $e_{SI}^\infty(k) = nk - e_{II}(k)$ összefüggést a rekurzió $e_{SI}^\infty(k)$ -ra felírható

$$n - (e_{SI}^\infty(k+1) - e_{SI}^\infty(k)) = e_{II}(k+1) - e_{II}(k) \quad (3.13)$$

formában. Vezessük be az $x = \frac{k}{N} \in [0, 1]$ jelölést, és legyen $f : [0, 1] \rightarrow \mathbb{R}$ egy olyan differenciálható függvény, amire fennáll, hogy $f(\frac{k}{N}) = e_{SI}^\infty(k)$. Ekkor

$$e_{SI}^\infty(k+1) - e_{SI}^\infty(k) = f\left(\frac{k+1}{N}\right) - f\left(\frac{k}{N}\right) \approx \frac{1}{N} f'\left(\frac{k}{N}\right),$$

és minél nagyobb N , annál pontosabb az utolsó közelítés. Így a (3.8) és (3.13) összefüggések a következő differenciálegyenlethez vezetnek f -re:

$$n - \frac{1}{N} f'(x) = 2 \frac{nN(1-x) + (n-1)f(x) - n}{nN(1-x) - 1},$$

felhasználva, hogy $m = nk - e_{II}(k) = f(x)$ és $n(N-k) = nN(1-x)$. Átrendezve az egyenletet az

$$\frac{1}{N} f'(x) + 2f(x) \frac{n-1}{nN(1-x) - 1} = \frac{(n-2)nN(1-x) + n}{nN(1-x) - 1}, \quad (3.14)$$

elsőrendű lineáris differenciálegyenletet kapjuk, ami könnyedén megoldható az f függvényre. Első lépésben megoldjuk a homogén egyenletet, majd az állandók variálásának módszerével megkapjuk a differenciálegyenlet megoldását. A megoldás

$$f(x) = (nN(1-x) - 1)^{2-2/n} K(x)$$

alakban keresendő, ahol K ismeretlen függvény. Ezt az alakot behelyettesítve a (3.14) egyenletbe, K' -ra a

$$-\frac{1}{N} K'(x) = (2-n) (nN(1-x) - 1)^{2/n-2} - (2n-2) (nN(1-x) - 1)^{2/n-3}.$$

összefüggést kapjuk, amit visszaintegrálva

$$K(x) = (nN(1-x) - 1)^{2/n-1} + (nN(1-x) - 1)^{2/n-2} + c$$

kifejezés jön ki, ahol $c \in \mathbb{R}$ konstans. Így

$$f(x) = nN(1-x) + c(nN(1-x) - 1)^{2-2/n}.$$

A c konstans az $f(\frac{1}{N}) = n$ kezdeti feltételből számíthatjuk ki, ami ekvivalens az $e_{SI}^\infty(1) = n$ összefüggéssel. Ezek alapján

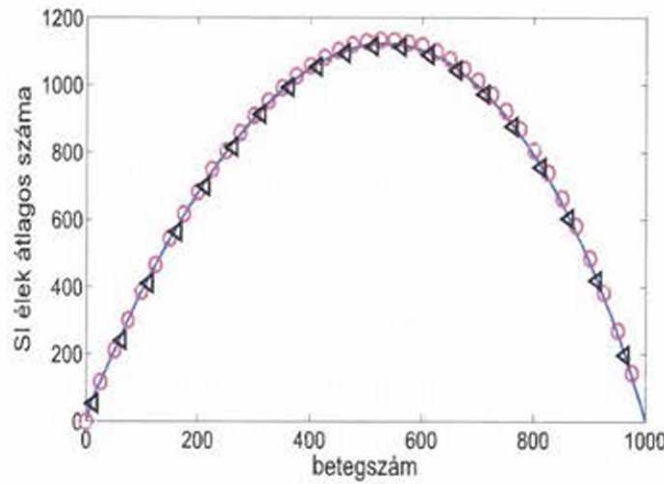
$$f(x) = nN(1-x) - n(N-2) \left(\frac{nN(1-x) - 1}{n(N-1) - 1} \right)^{2-2/n}. \quad (3.15)$$

Felhasználva, hogy $e_{SI}^\infty(k) = f(\frac{k}{N})$, megkapjuk az explicit kifejezést $e_{SI}^\infty(k)$ -ra:

$$e_{SI}^\infty(k) = nN \left(1 - \frac{k}{N}\right) - n(N-2) \left(\frac{nN(1 - \frac{k}{N}) - 1}{n(N-1) - 1} \right)^{2-2/n}.$$

□

A 3.6. ábrán összevetjük $e_{SI}^\infty(k)$ értékeit a szimulációból, majd a (3.8) rekurzióból kiszámítva illetve a most ismertetett (3.12) explicit formula alapján megadva. Nyilván N értékét minél nagyobbra érdemes választani, hogy az explicit kifejezés jól közelítse a rekurziós képletből kapott eredményt.



3.6. ábra. A $(k, e_{SI}(k))$ görbék a szimulációból (\circ), a rekurzióból $e_{SI}^\infty(k) = nk - e_{II}(k)$ alapján (kék görbe) és a (3.14) differenciálegyenlet megoldása (\triangleleft) reguláris random gráf esetén $N = 1000$, $\tau = 10$, $\gamma = 1$, $n = 6$. 250 szimuláció átlagából, 10 kezdeti beteggel.

3.4.2. Az SI élek maximális száma τ függvényében

Most mutatunk egy elméleti módszert arra, hogy hogyan lehet kiszámítani $e_{SI}(k)$ értékeit kisebb τ esetén a rekurzióból kapott $e_{SI}^\infty(k)$ értékekből, avagy más τ/γ arány mellett. A 3.5. ábrán látható, hogy a τ értékének a megválasztása befolyásolja a $(k, e_{SI}(k))$ ($k = 0, 1, \dots, N$) pontok által meghatározott görbének a maximum értékét, viszont azt fel fogjuk tenni ezzel ellentétben, hogy a maximum hely minden τ esetén változatlan, ugyanis az ábra alapján lényegében ez állandónak tekinthető. Erre a megállapításra azért van szükség, hogy a maximum helyet közvetlenül megkapjuk a nagy τ esetén előállított $(k, e_{SI}^\infty(k))$ értékekből. Az SI élek maximum

értékének a kiszámítására, melyet jelöljünk $[SI]_{max}$ -al, mutatunk itt egy eljárást. Ehhez a (2.9) egyenleteket hívjuk segítségül, amit a megfelelő γ és τ értékekkel írunk fel:

$$\begin{aligned}\dot{[I]} &= \tau[SI] - \gamma[I], \\ \dot{[SI]} &= \gamma([II] - [SI]) + \tau([SSI] - [ISI] - [SI]), \\ \dot{[II]} &= -2\gamma[II] + 2\tau([ISI] + [SI]), \\ \dot{[SS]} &= 2\gamma[SI] - 2\tau[SSI],\end{aligned}$$

és a szokásos lezárást alkalmazzuk:

$$[SSI] \simeq \frac{n-1}{n} \frac{[SS][SI]}{[S]}, \quad [ISI] \simeq \frac{n-1}{n} \frac{[SI]^2}{[S]}.$$

Az SI élek száma akkor lehet maximális, ha $\dot{[SI]} = 0$, azaz ha

$$\gamma([II] - [SI]) = \tau([SI] + [ISI] - [SSI]).$$

Felhasználva a

$$[II] + [SI] = n[I], \quad [SS] + [SI] = n[S]$$

formulákat és a lezárási képleteket, fennáll a következő összefüggés:

$$\gamma n[S](n[I] - 2[SI]) = \tau[SI](n[S] + (n-1)[SI] - (n-1)(n[S] - [SI])).$$

Mivel $[S] = N - [I]$, egy másodfokú egyenletet kapunk az $[SI]$ -re

$$2\tau(n-1)[SI]^2 + n(N-[I])(2\tau - n\tau + 2\gamma)[SI] - n^2\gamma(N-[I])[I] = 0. \quad (3.16)$$

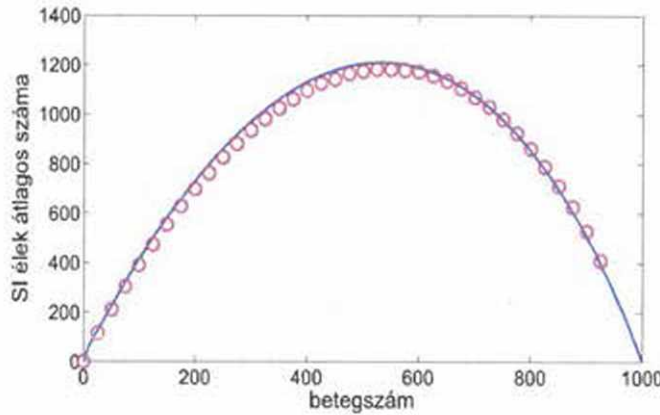
Ennek a pozitív megoldása adja meg az SI élek maximális értékét, vagyis $[SI]_{max}$ -ot. Az egyenletben $[I]$ jelöli a $(k, e_{SI}^\infty(k))$ ($k = 0, 1, \dots, N$) pontok által megadott görbének a maximum helyét, tehát először ezt kell meghatároznunk. Ezt többféle módon tehetjük meg. Egyrészt megnézhetjük, hogy melyik k esetén a legnagyobb $e_{SI}^\infty(k)$ értéke, jelöljük ezt k_{max} -al, majd $[I]_{max} := k_{max}$. Másrészt reguláris véletlen gráf esetén az előbbinél pontosabban, analitikusan is kiszámolhatjuk az előző részben megadott differenciálható $f : [0, 1] \rightarrow \mathbb{R}$ függvény maximum helyét: legyen ez x_{max} , majd $[I]_{max} := Nx_{max}$. A (3.16) egyenletben az $[I]$ -k helyére be kell írni $[I]_{max}$ értékét, majd a másodfokú egyenlet pozitív megoldása lesz a keresett $[SI]_{max}$.

Ha megvan $[I]_{max}$ és $[SI]_{max}$ is adott τ esetén, akkor az $e_{SI}(k)$ értékeket úgy kaphatjuk meg a $(k, e_{SI}^\infty(k))$ ($k = 0, 1, \dots, N$) pontok által meghatározott parabolyszerű görbéből, hogy az e_{SI}^∞

görbét átranszformáljuk úgy, hogy az $([I]_{max}, [SI]_{max}^{\infty})$ pont feltolódjon az $([I]_{max}, [SI]_{max})$ pontba, ahol $[SI]_{max}^{\infty}$ jelöli az e_{SI}^{∞} görbe maximum értékét, vagyis $[SI]_{max}^{\infty} = e_{SI}^{\infty}(k_{max})$, illetve reguláris véletlen gráf esetén $[SI]_{max}^{\infty} = f(x_{max})$. Precízen leírva

$$e_{SI}(k) := e_{SI}^{\infty}(k) \frac{[SI]_{max}}{[SI]_{max}^{\infty}}. \quad (3.17)$$

Ennek a módszernek az eredményét hasonlítottuk össze a szimulációval a 3.7. ábrán reguláris véletlen gráf esetén, ami láthatóan jó eredményt ad. Azonban heterogén fokszámeloszlású gráfon a (2.9) egyenletek segítségével visszatranszformált e_{SI} görbe nem adja a valós értékek egy jó közelítését. Ebben az esetben érdemes lehet a 5.3.1. alszakaszban ismertetésre kerülő (5.9) egyenleteket használni az e_{SI} görbe maximum értékének meghatározására.



3.7. ábra. A $(k, e_{SI}(k))$ görbék a szimulációból (\circ) illetve a rekurzióból (kék görbe) a (3.17) képlet segítségével reguláris véletlen gráf esetén $N = 1000$, $\tau = 2$, $\gamma = 1$, $n = 6$. 250 szimuláció átlagából, 10 kezdeti beteggel.

3.5. Az eljárás áttekintése

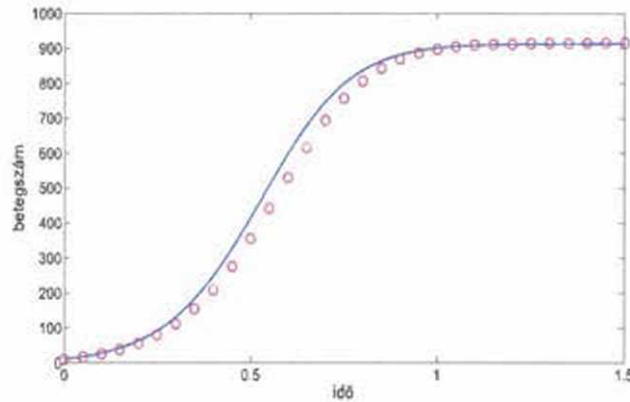
Összegzésül listába szedjük az analitikus módszerünk főbb lépéseit. Vesszünk egy tetszőleges konfigurációs modellel generált véletlen gráfot, aminek ismerjük a fokszámeloszlását, emellett legyen adott τ és γ is. Meg kell határoznunk a (3.3) egyenletrendszer a_k és c_k együtthatóit, de mivel $c_k = \gamma k$, ezért csak az a_k értékek kiszámításával kell részletesebben foglalkoznunk. Az algoritmus a következő:

1. A foksámeloszlás alapján kiszámítjuk az II élek átlagos számát a betegszám függvényében $\tau \rightarrow +\infty$ esetén, azaz $e_{II}(k)$ -t, felhasználva a (3.11) rekurziós összefüggést, lásd 3.3. szakasz. (Megjegyezzük, hogy ez az érték még független az előre megadott τ -tól.)
2. Ez alapján megadjuk az SI élek átlagos számát a betegszám függvényében $\tau \rightarrow +\infty$ esetén a $e_{SI}^\infty(k) = \sum_{l=1}^L n_l I_l(k) - e_{II}(k)$ formula segítségével, aminek leírása a 3.4. szakaszban található.
3. Az SI élek átlagos száma, azaz e_{SI} adott γ és τ értékek esetén meghatározható a (3.17) összefüggés segítségével, lásd 3.4.2. alszakasz. Majd a keresett együtthatók: $a_k = \tau e_{SI}(k)$, $k = 0, 1, \dots, N$.
4. Felhasználva az elmélet alapján kiszámolt a_k és c_k együtthatókat a (3.3) redukált egyenlet numerikusan megoldható és az átlagos betegszámot, amit az $I(t) = \sum_{k=0}^N k x_k(t)$ formulával kapunk, összehasonlíthatjuk a szimulációs eredményekkel.

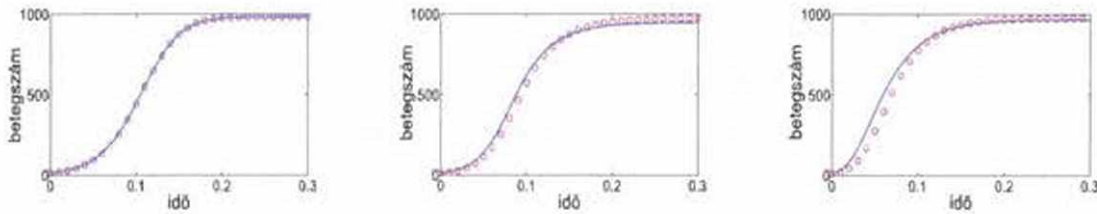
Az eljárás szemléltetéseként végrehajtottuk az algoritmust egy reguláris véletlen gráfon kis τ mellett, illetve egy reguláris véletlen, egy bimodális és egy Barabás-Albert gráfon nagy τ esetén. Az eredményeket összevetettük a szimulációból kimért betegszám aránnyal, amit a 3.8. és 3.9 ábrákon prezentálunk. Látható, hogy a kínált módszer nagyon jó közelítést adja a szimulációs eredményeknek, ami azt mutatja, hogy az állapottér ilyen drasztikus lecsökkentése mellett is van esély a dinamika legfontosabb jellemzőinek megőrzésére, illetve a redukált egyenletrendszer együtthatóiba sikeresen belekódolhatóak a gráfstruktúra azon speciális tulajdonságai, amik a járványterjedés eltérő lefolyását eredményezik a különböző gráftípusokon.

3.6. Összefoglalás

Ebben a fejezetben formalizáltunk egy új típusú közelítő differenciálegyenlet-rendszert SIS típusú dinamika mellett tetszőleges foksámeloszlású konfigurációs gráf esetén. Az ötletet a teljes gráf esetén alkalmazható, pontos, alacsony dimenziós differenciálegyenlet-rendszer inspirálta. Kiderült, hogy a rendszer redukciójának érdekében bevetésre kerülő $N + 1$ eleműre csökkentett állapottér használata hatékony eljárásához vezet és az átmeneti ráták kombinatorikus és analitikus eszközökkel kiszámíthatóak. Jelenleg a kalkulációk csak a konfigurációs



3.8. ábra. Az átlagos betegszám az idő függvényében a szimulációból (\circ) és a (3.3) redukált egyenletekből (kék görbe) reguláris véletlen gráf esetén, $N = 1000$, $n = 6$, $\tau = 2$, $\gamma = 1$. 250 szimuláció átlagából, 10 kezdeti beteggel.



3.9. ábra. Az átlagos betegszám az idő függvényében a szimulációból (\circ) és a (3.3) redukált egyenletekből (kék görbe) rendre reguláris véletlen gráfon, bimodális véletlen gráfon és Barabási-Albert gráfon, $N = 1000$, $n = 6$, $\tau = 10$, $\gamma = 1$. 250 szimuláció átlagából, 10 kezdeti beteggel.

modellel generált gráfokra érvényesek, de a numerikus eredmények azt mutatják, hogy az eljárás kiterjeszthető klaszterezett gráfokra is, lásd 3.1. ábra, habár a fertőzési ráták kiszámítása bonyolultabbá válhat.

Érdekes észrevétel, hogy az átmeneti ráták valójában valószínűségi változók is lehetnének, amik megadják az SI élek számát rögzített betegszám mellett, tekintve, hogy a k beteg sok különböző konfigurációban jelenhet meg a gráfon. A jelenlegi modellben csupán az SI élek várható értékének meghatározására koncentráltunk, azonban ez a gondolat azt az ötletet adja, hogy a modell egy kiterjesztése lehetne egy olyan differenciálegyenlet-rendszer család vizsgálata, ahol az átmeneti ráták valamilyen előre meghatározott eloszlású valószínűségi változók.

Az eljárás során kiindulunk a gráf fokszámeloszlásából és ennek alapján kiszámítjuk az

átmeneti rátákat. Mivel az átmeneti ráták görbéje minden gráftípus esetén nagyon speciális alakú, ezért ezen görbék egyfajta lenyomatai az adott gráfstruktúra jellemzőinek. Emiatt felmerül a kérdés, hogy visszafelé is lehet-e gondolkodni, azaz az e_{SI} görbékből vissza lehet-e következtetni a hálózat néhány fontos tulajdonságára, akár például a fokszámoszlásra. Ez a megfigyelés további kutatások tárgya lehet a későbbiekben.

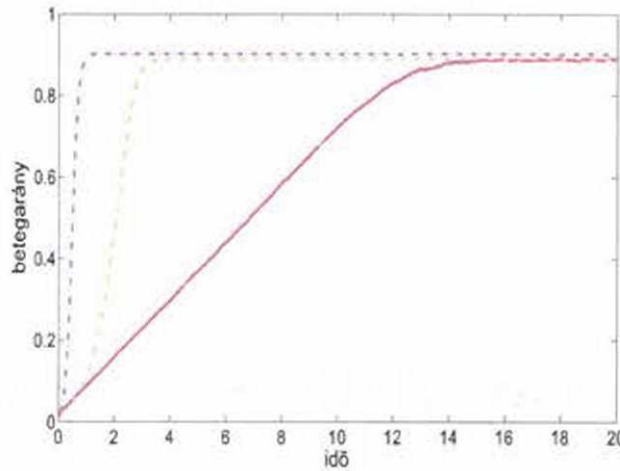
A 3.4.1. alszakaszban explicit képletet adtunk az SI élek átlagos számára reguláris véletlen gráf esetén, lásd (3.15). A leggyakrabban használt képlet a jól ismert $[SI] = n[I] \frac{N-[I]}{N-1}$, aminek a megfelelője függvény alakban $f(x) = \frac{nN^2}{N-1}x(1-x)$, ahol $x = [I]/N$. Mivel az előbbi (3.7) becslés a járványterjedés folyamata alatt végbemenő korrelációkat figyelmen kívül hagyja, egy általánosabb függvényt ajánlunk az SI élek átlagos számának közelítésére, nevezetesen az $f(x) = kx^p(1-x)^q$ alakot, valamilyen alkalmas k , p és q paraméterek mellett [24]. A (3.15) alak egy kicsit más, de cserébe analitikusan levezethető, viszont numerikusan nagy N -re a (3.15) függvény és a most ajánlott, általánosított függvény közelítőleg megegyezik. Az e_{SI} görbe $f(x) = kx^p(1-x)^q$ alakban való közelítő meghatározása is egy jövőbeli kutatási irány lehet, ahol a paraméterek megválasztását a gráfstruktúra befolyásolja.

4. fejezet

Betegségterjedés modellezése speciális struktúrájú gráf esetén

4.1. Motiváció

Az előző fejezetben adtunk egy jól működő módszert járványterjedés modellezésére *SIS* dinamika mellett általános foksámeloszlású konfigurációs gráf esetén, aminek egy speciális esete a reguláris véletlen gráf. Felmerül a kérdés, hogy ha a reguláris véletlen gráf átlagos fokszáma 2, azaz egy körgráfon vizsgálunk járványterjedést, akkor a gráf speciális struktúrából fakadóan egyéb előnyökhöz is juthatunk-e, vagyis más kedvező eljárásokat is ki lehet-e dolgozni ilyenkor. Ráadásul ezen gráftípus esetén az alacsony dimenziós közelítő módszerek, mint a (2.7) és a (2.9) egyenletek, kifejezetten rossz eredményt adnak (lásd 4.1. ábra), ugyanis ebben az esetben a betegség frontszerűen halad végig a gráfon. Ebben a fejezetben három módszert adunk a betegség elterjedésének modellezésére körgráfon, azaz a betegség eltűnésének esetével nem foglalkozunk. Az első eljárásban az eredeti Markov-folyamathoz kapcsolódó alapegyenletekből kiindulva felírunk egy új differenciálegyenlet-rendszert a betegek száma alapján, ezen az eljáráson belül kétféle technikát is alkalmazunk. Ekkor az *SI* típusú élek átlagos számának újfajta meghatározása jelent kihívást. A második esetben egy új Markov-lánc bevezetésével az előbb említett betegségfront hosszának változását követjük nyomon. A harmadik módszernél az élek különböző állapotaira írunk fel differenciálegyenleteket. Mindhárom esetben a betegek átlagos számát becsüljük meg az idő függvényében.



4.1. ábra. A betegek arányának várható értéke az idő függvényében körgráfon a szimulációból (piros görbe) a (2.7) egyenletből (lila szaggatott görbe) és a (2.9) rendszerből (barna szaggatott görbe) körgráf esetén. $N = 100$, $\gamma = 1$, $\tau = 5$. 500 szimuláció átlagából, 1 kezdeti beteggel.

4.2. Redukált alapegyenletek a betegszám függvényében

Ebben a részben az N csúcsú körgráfot helyezzük a figyelem középpontjába és SIS típusú járványterjést vizsgálunk rajta. Ugyanúgy, mint az előző fejezetben, a (2.2) alapegyenletek számát fogjuk 2^N eleműről $N + 1$ -re csökkenteni, úgy hogy az új állapottér a $\{0, 1, \dots, N\}$ halmaz legyen. A redukált egyenletek

$$\dot{x}_k(t) = a_{k-1}x_{k-1}(t) - (a_k + c_k)x_k(t) + c_{k+1}x_{k+1}(t), \quad k = 0, 1, \dots, N \quad (4.1)$$

alakúak, ahol $x_k(t)$ jelöli - szokásos módon - annak a valószínűségét, hogy k beteg csúcs található a gráfban a t időpontban. Emellett egyrészt $c_k = \gamma k$ a gyógyulásért felelős együttható, másrészt $a_k = \tau e_{SI}(k)$, ahol $e_{SI}(k)$ jelöli az SI élek valamilyen módon meghatározott átlagos számát k beteg esetén. Felidézve a 3.2. szakaszban írtakat, a differenciálegyenletet a

$$\dot{\mathbf{x}}_k = \tau \cdot N_{SI}(\mathcal{S}^{k-1})\mathbf{x}_{k-1} + B_k\mathbf{x}_k + (k+1)\gamma\mathbf{x}_{k+1}$$

alakban is vizsgálhatjuk, ahol

$$\mathbf{x}_k = (x_{S_1^k}, x_{S_2^k}, \dots, x_{S_d^k})^T, \quad k = 0, \dots, N,$$

$$N_{SI}(\mathcal{S}^{k-1}) = (N_{SI}(\mathcal{S}_1^{k-1}), N_{SI}(\mathcal{S}_2^{k-1}), \dots, N_{SI}(\mathcal{S}_{d_k}^{k-1}))$$

és $N_{SI}(\mathcal{S}_j^{k-1})$ az \mathcal{S}_j^{k-1} állapotban lévő SI élek száma. Az alszakasz célkitűzése olyan a_k együttható felírása, melyre

$$\tau \cdot N_{SI}(\mathcal{S}^k) x_k = a_k x_k.$$

A problémát az okozza, hogy az $N_{SI}(\mathcal{S}^k)$ vektor elemei különbözőek, hiszen a k beteget tartalmazó különböző állapotokban más lehet az SI élek száma, így az \mathcal{S}^k -beli tagoknak a szorzói nem azonosak. Ezért az $N_{SI}(\mathcal{S}^k)$ sorvektor helyett egy olyan vektort szeretnénk használni, amelynek minden koordinátája megegyezik, így ki lehetne emelni, azaz egy

$$N_{SI}(\mathcal{S}^k) \approx e_{SI}(k) e_k, \quad (4.2)$$

becsléssel szeretnénk élni, ahol $e_{SI}(k)$ jelöli az SI élek egyfajta várható értékét. Ekkor a_k együtthatóra teljesülne, hogy

$$a_k = \tau e_{SI}(k). \quad (4.3)$$

A továbbiakban két új lehetőséget ajánlunk az $e_{SI}(k)$ értékek meghatározására körgráf esetén.

4.2.1. Becslés az SI élek átlagos számára az állapotok súlyozásával

Most egy új módszerrel próbálkozunk, miszerint legyen $e_{SI}(k)$ az $N_{SI}(\mathcal{S}^k)$ vektor egy súlyozott átlaga, azaz

$$e_{SI}(k) = \sum_{j=1}^{d_k} N_{SI}(\mathcal{S}_j^k) \cdot w_{kj}, \quad k = 1, \dots, N, \quad (4.4)$$

ahol w_k a súlyokat tartalmazó, d_k dimenziós vektor, és a súlyok összege 1, azaz $\sum_{j=1}^{d_k} w_{kj} = 1$. Nyilván $e_{SI}(0) = 0$.

A legegyszerűbb választás, ha minden súlyt egyenlőnek veszünk, azaz $w_{kj} = 1/d_k$, $j = 1, 2, \dots, d_k$. Reguláris véletlen gráfon az egyenletes súlyozás a (3.7) becsléshez vezet, amit egyszerű kombinatorikai érveléssel be lehet látni.

4. Lemma. *Tekintsünk egy reguláris véletlen gráfot, azaz legyen minden csúcs n fokú. Ha a súlyok egyenlőek, vagyis $w_{kj} = 1/d_k$, akkor*

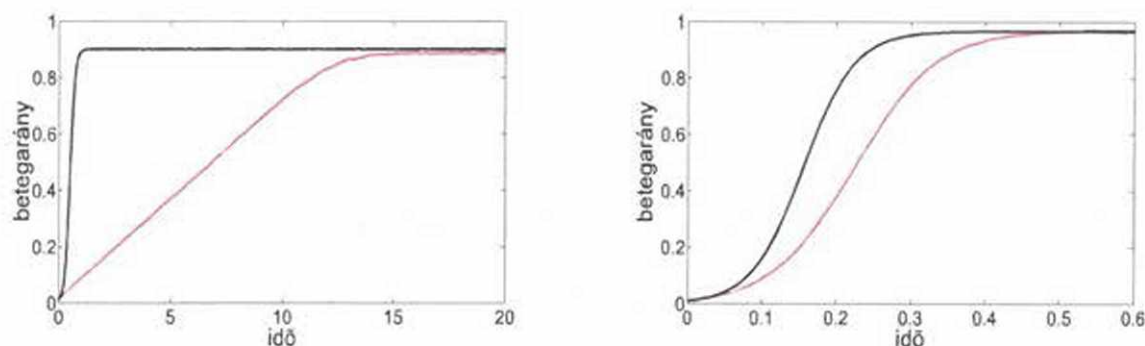
$$e_{SI}(k) = \frac{1}{d_k} \cdot \sum_{j=1}^{d_k} N_{SI}(\mathcal{S}_j^k) = \frac{n}{N-1} \cdot k \cdot (N-k). \quad (4.5)$$

BIZONYÍTÁS. Tegyük fel, hogy k beteg van a gráfban. Vegyünk egy rögzített élet és vizsgáljuk meg, hogy hányféle konfigurációban lesz SI típusú ez az él. Az él egyik csúcsa legyen beteg, a

másik egészséges, majd a maradék $k-1$ beteget szétszórjuk a további $N-2$ csúcson, azaz az S^k halmaz $2 \cdot \binom{N-2}{k-1}$ különböző eleme esetén lesz a kiválasztott él SI típusú, hiszen a kiválasztott élen kétféleképpen helyezkedhet el a beteg és az egészséges csúcs. Tehát egy konkrét élet a $\sum_{j=1}^{d_k} N_{SI}(S_j^k)$ összegbe $2 \cdot \binom{N-2}{k-1}$ -szor számolunk bele. Mivel reguláris véletlen gráfon az összes él száma $\frac{nN}{2}$, ezért az összes élet $\frac{nN}{2} \cdot 2 \cdot \binom{N-2}{k-1}$ -szor vesszük figyelembe az összegzés során. Tehát

$$\begin{aligned} \sum_{j=1}^{d_k} N_{SI}(S_j^k) &= nN \cdot \binom{N-2}{k-1} = \frac{n}{N-1} \cdot N(N-1) \cdot \frac{(N-2)!}{(k-1)!(N-k-1)!} = \\ &= \frac{n}{N-1} \cdot k(N-k) \cdot \frac{N!}{k!(N-k)!} = \frac{n}{N-1} \cdot k(N-k) \cdot d_k. \end{aligned}$$

□



4.2. ábra. A betegek arányának várható értéke az idő függvényében a szimulációból (piros görbe) és a (4.5) képletből kiszámolt (4.1) redukált egyenletrendszerből (lila szaggatott görbe) körgráf esetén (baloldali ábra) és reguláris véletlen gráf esetén, ahol minden csúcs foka 6 (jobb oldali ábra). $N = 100$, $\gamma = 1$, $\tau = 5$. 500 szimuláció átlagából, 1 kezdeti beteggel.

Minél nagyobb a reguláris véletlen gráf átlagos foka, annál jobb egyezést mutat a szimulációval az egyenletes súlyozás, amit a 4.2. ábrán szemléltetünk. Teljes gráf esetén, amikor $n = N - 1$, az egyezés tökéletes, azonban körgráf esetén kifejezetten rossz közelítést ad, ami azt mutatja, hogy az egyenletes súlyozással megadott redukált egyenletek megoldása nem tudja visszatükrözni a valódi folyamatot. Ennek az az oka, hogy a betegségterjedés során sokkal valószínűbb a megjelenése az S^k halmaz azon állapotainak, ahol a betegek egymás mellett csoportosulnak, mint azon az állapotoknak, ahol a betegek szétszórva helyezkednek el a gráfon.

A továbbiakban az alszakaszban két állítást fogalmazunk meg, majd megvizsgáljuk a körgráfon való alkalmazhatóságukat.

Megint tekintsük a 2^N dimenziós (2.2) alapegyenlet-rendszert:

$$\dot{x}(t) = Px(t),$$

ahol a P mátrix természetesen a 2.3. alszakaszban ismerttetett blokk tridiagonális alakú. Ismeretes, hogy mivel P egy sztochasztikus mátrix, ezért az első sajátértéke: $\lambda_0 = 0$ és a hozzá tartozó sajátvektor $(1, 0, \dots, 0)^T$. Ez a differenciálegyenlet stacionárius állapotának felel meg, illetve a Markov-folyamat nyelő állapotának, ebben az állapotban minden csúcs egészséges. A P mátrix többi sajátértéke negatív valósrésztű. A továbbiakban hívjuk második sajátértéknek a legnagyobb nem nulla valósrésztűt, jelöljük λ_1 -el és a hozzá tartozó sajátvektort v_1 -el, ami megadja az ún. kvázi-stacionárius állapotot. Könnyen látható, hogy a rendszer megoldásai az egyensúlyi állapothoz konvergálnak, azonban az állapotok nagyon lassan kerülnek a teljesen egészséges állapot közelébe, előtte sokáig időznek a kvázi-stacionárius állapotban, ugyanis λ_1 nagyon közel van a nullához [20]. Emiatt a v_1 sajátvektor fontos szerepet kap a rendszer viselkedésének tanulmányozása során. Az ötlet az, hogy az állapotok súlyozását a kvázi-egyensúlyi állapothoz tartozó v_1 sajátvektor alapján végezzük.

Ahelyett, hogy kiszámítanánk a v_1 sajátvektort, ami lényegében kivitelezhetetlen egy ekkora méretű mátrix esetén, a

$$\tilde{P} = \begin{pmatrix} B_1 & C_2 & 0 & 0 & 0 \\ A_1 & B_2 & C_3 & 0 & 0 \\ 0 & A_2 & B_3 & C_4 & 0 \\ \vdots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & A_{N-1} & B_N \end{pmatrix}$$

mátrix 0-hoz tartozó sajátvektorával közelítjük, ahol a \tilde{P} mátrixot úgy kapjuk, hogy a P mátrix első sorát és oszlopát elhagyjuk, és a B_1 mátrixot úgy módosítjuk, hogy megint egy sztochasztikus mátrixot kapjuk, tehát az új B_1 legyen egy diagonális mátrix a főátlójában a $-e_2 A_1$ vektorral. Ezt az eljárást úgy is interpretálhatjuk, hogy az \mathcal{S} állapottérből kihagyjuk az \mathcal{S}^0 állapotot, azaz azt az esetet, amikor minden csúcs egészséges. Ekkor ennek az új Markov-folyamatnak az egyensúlyi állapota adja az eredeti folyamat kvázi-stacionárius állapotát.

5. Állítás. Tetszőleges gráf esetén legyen u a \tilde{P} mátrix 0-hoz tartozó sajátvektora és u_k az állapottér k beteget tartalmazó állapotaihoz tartozó része u -nak. Tehát $u = (u_1, u_2, \dots, u_N)^T$

és u_k egy d_k dimenziós vektor. A w_k súlyokat határozzuk meg az u_k vektor alapján, vagyis

$$w_k := \frac{u_k}{e_k u_k}, \quad k = 1, \dots, N. \quad (4.6)$$

Ha $e_{SI}(k)$ a (4.4) formulával van megadva, $k = 1, \dots, N$, akkor minden $k = 1, \dots, N - 1$ esetén

$$e_{SI}(k) = (k+1) \cdot \frac{\gamma}{\tau} \cdot \frac{e_{k+1} u_{k+1}}{e_k u_k}.$$

BIZONYÍTÁS. Mivel u a \tilde{P} mátrix 0-hoz tartozó sajátvektora, azért a $\tilde{P}u = 0$ összefüggést részletesen kiírva

$$\begin{aligned} B_1 u_1 + C_2 u_2 &= 0, \\ A_1 u_1 + B_2 u_2 + C_3 u_3 &= 0, \\ A_2 u_2 + B_3 u_3 + C_4 u_4 &= 0, \\ &\vdots \\ A_{N-1} u_{N-1} + B_N u_N &= 0. \end{aligned}$$

A \tilde{P} mátrix oszlopainak összege 0, azaz $e_k B_k = -(e_{k-1} C_k + e_{k+1} A_k)$, $k = 2, \dots, N - 1$, emellett $e_1 B_1 = -e_2 A_1$ és $e_N B_N = -e_{N-1} C_N$ is fennáll $k = 1$ és $k = N$ esetén. Így ha az előbbi egyenleteket balról megszorozzuk az e_k sorvektorokkal és $e_k B_k$ helyére behelyettesítjük az említett formulát, akkor a következő N egyenletből álló összefüggő rendszer adódik:

$$\begin{aligned} -e_2 A_1 u_1 + e_1 C_2 u_2 &= 0, \\ e_2 A_1 u_1 - (e_1 C_2 + e_3 A_2) u_2 + e_2 C_3 u_3 &= 0, \\ &\vdots \\ e_N A_{N-1} u_{N-1} - e_{N-1} C_N u_N &= 0. \end{aligned}$$

Az első k egyenletet összegezve

$$e_k C_{k+1} u_{k+1} = e_{k+1} A_k u_k, \quad k = 1, \dots, N.$$

Most felhasználjuk az ismert $e_{k-1} C_k = k \gamma e_k$ és $N_{SI}(S^k) = \frac{1}{\tau} \cdot e_{k+1} A_k$ formulákat. Az első alkalmazva

$$(k+1) \gamma e_{k+1} u_{k+1} = e_{k+1} A_k u_k,$$

majd a másodikat és a (4.4) formulát, illetve w_k definícióját:

$$e_{SI}(k) = N_{SI}(S^k) \cdot w_k = \frac{1}{\tau} \cdot e_{k+1} A_k \cdot \frac{u_k}{e_k u_k} = \frac{1}{\tau} \cdot \frac{(k+1) \gamma e_{k+1} u_{k+1}}{e_k u_k} = (k+1) \cdot \frac{\gamma}{\tau} \cdot \frac{e_{k+1} u_{k+1}}{e_k u_k}.$$

□

Felírva $N = 5$ esetén az előbbi állítást körgráfon, a következőt kapjuk:

$$\begin{aligned} e_{SI}(1) &= 2, \\ e_{SI}(2) &= 3 \cdot \frac{\gamma}{\tau} \cdot \frac{2\tau(3\gamma^2 + 8\tau\gamma + 6\tau^2)}{\gamma(9\gamma^2 + 21\tau\gamma + 14\tau^2)}, \\ e_{SI}(3) &= 4 \cdot \frac{\gamma}{\tau} \cdot \frac{\tau(3\gamma^2 + 10\tau\gamma + 8\tau^2)}{2\gamma(3\gamma^2 + 8\tau\gamma + 6\tau^2)}, \\ e_{SI}(4) &= 2. \end{aligned}$$

Behelyettesítve a (4.3) formulába, megkapjuk a (4.1) egyenletrendszer együtthatóit.

Bebizonyítható, - ahogy $N = 5$ esetén is látható - hogy tetszőleges méretű körgráf esetén az $e_{SI}(k)$ értékek racionális törtfüggvény formában előállíthatóak, ezzel foglalkozunk a következő állításban.

6. Állítás. *Körgráf esetén legyen u a \tilde{P} mátrix 0-hoz tartozó sajátvektora. Ha az $e_{SI}(k)$ értékek a (4.4) formulával és a (4.6) súlyozással vannak kiszámítva, akkor az $e_{SI}(k)$ kifejezések racionális törtfüggvény alakban megadhatóak, azaz*

$$e_{SI}(k) = (k+1) \cdot \frac{1}{r} \cdot \frac{p_k(r)}{q_k(r)},$$

alkalmas $p_k, q_k \in \mathbb{R}[r]$ polinomokkal, ahol $r = \frac{\tau}{\gamma}$, tehát $e_{SI}(k)$ előáll r polinomjainak hányadosaként, $k = 0, \dots, N$.

BIZONYÍTÁS. Az alap feltevésünk, hogy u a 0-hoz tartozó sajátvektor, azaz $\tilde{P}u = 0$. Az u vektor koordinátáiról szeretnénk valamilyen információhoz jutni. Legyen \tilde{T} az állapotokhoz tartozó $\tilde{K} \times 2^N - 1$ dimenziós összevonás mátrix, ahol \tilde{K} az állapotosztályok száma. Azaz $\tilde{T}_{i,j} = 1$ pontosan akkor, ha a j -edik állapot az i -edik összevont állapotosztályban található, egyébként $\tilde{T}_{i,j} = 0$. Keressük az u vektort $u = \tilde{T}^T \tilde{u}$ alakban, ugyanis az összevonható állapotoknak ugyanakkora súlyt szeretnénk adni és a \tilde{K} dimenziós \tilde{u} vektorban tároljuk az egyes összevont állapotosztályok súlyait. Tehát a

$$\tilde{P}\tilde{T}^T \tilde{u} = 0 \quad (4.7)$$

túlhatározott lineáris egyenletrendszer megoldásáról szeretnénk információt szerezni. A (4.7) rendszer helyett azonban a

$$\tilde{T}\tilde{P}\tilde{T}^T \tilde{u} = 0 \quad (4.8)$$

rendszert tekintjük, ugyanis ennek együttható mátrixa $\tilde{T}\tilde{P}\tilde{T}^T \in \mathbb{R}^{\tilde{K} \times \tilde{K}}$ négyzetes, így könnyebben vizsgálható és a megoldása megegyezik a (4.8) rendszer megoldásával. Ennek belátásához vegyük az $\tilde{y} = \tilde{B}\tilde{y}$ összevont rendszer \tilde{B} együttható mátrixát, ahogy a 2.4.2. szakaszban ismertettük. Erről tudjuk, hogy $\tilde{T}\tilde{P} = \tilde{B}\tilde{T}$ fennáll, azaz $\tilde{T}\tilde{P}\tilde{T}^T = \tilde{B}\tilde{T}\tilde{T}^T$ is. Vagyis a (4.8) rendszer ekvivalens a $\tilde{B}\tilde{T}\tilde{T}^T\tilde{u} = 0$ rendszerrel. Könnyen látható, hogy a $\tilde{T}\tilde{T}^T$ valós értékű mátrix diagonális, így ezen rendszer pedig ekvivalens a

$$\tilde{B}\tilde{u} = 0 \quad (4.9)$$

rendszerrel. Mivel a \tilde{B} összevonással kapott együttható mátrix szintén sztochasztikus, ezért a (4.9) rendszernek létezik egy olyan egyértelmű \tilde{u} megoldása, melyre $u = \tilde{T}^T\tilde{u}$ alakban előáll.

Térjünk vissza a (4.8) lineáris egyenletrendszer megoldásának a vizsgálatára. A \tilde{P} mátrix felbontható egy csak τ -tól és egy csak γ -tól függő két mátrix összegére a következő módon:

$$\tilde{P} = \tau\tilde{P}_\tau + \gamma\tilde{P}_\gamma.$$

Ekkor $\tau F + \gamma G := \tilde{T}\tilde{P}\tilde{T}^T$, ahol $F = \tilde{T}\tilde{P}_\tau\tilde{T}^T$ és $G = \tilde{T}\tilde{P}_\gamma\tilde{T}^T$. Tehát \tilde{u} a $(\tau F + \gamma G)\tilde{u} = 0$ egyenletrendszer megoldása. γ -val osztva és bevezetve az $r = \frac{\tau}{\gamma}$ jelölést:

$$(rF + G)\tilde{u} = 0.$$

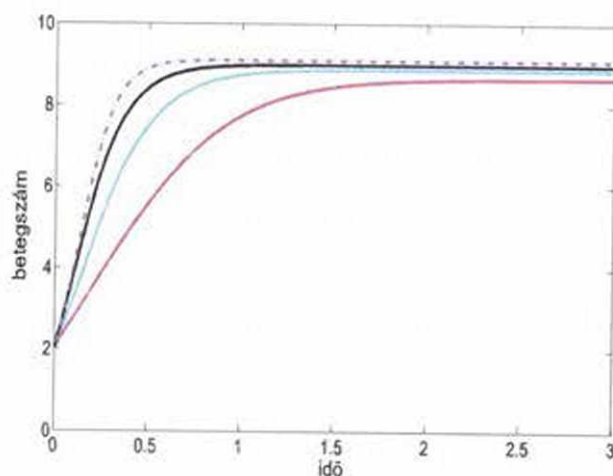
Mivel $rF + G$ egy $\tilde{K} \times \tilde{K}$ dimenziós sztochasztikus mátrix, így $\det(rF + G) = 0$, azonban ha elhagyjuk az utolsó sorát és (\tilde{u} utolsó koordinátáját 1-nek lerögzítve) a mátrix utolsó oszlopának (-1) -szeresét átvisszük az egyenletrendszer jobb oldalára, akkor alkalmazható az új - már $\tilde{K} - 1$ egyenletből álló - független rendszerre a Cramer-szabály.

Így kiszámítva az utolsó koordinátájától megfosztott \tilde{u} vektor koordinátáit, látható, hogy a vektor elemei $\frac{\tilde{p}(r)}{\tilde{q}(r)}$ alakúak, ahol $\tilde{p}, \tilde{q} \in \mathbb{R}[r]$, ugyanis $\tilde{u}_j = \frac{\det(r\tilde{F}_j + \tilde{G}_j)}{\det(rF + G)}$, ahol \tilde{F} és \tilde{G} az új egyenletrendszer mátrixai és \tilde{F}_j, \tilde{G}_j azon mátrixok, melyeket úgy kapunk, hogy az \tilde{F} illetve a \tilde{G} mátrixok j -edik oszlopát a jobboldali vektor r -hez tartozó illetve konstans komponenseire cseréljük a Cramer-szabály alapján.

Mivel $u = \tilde{T}^T\tilde{u}$ alakban adható meg, így $e_k u_k$ a \tilde{u} vektor bizonyos koordinátáinak összegeként áll elő és racionális törtfüggvények összege racionális törtfüggvény. Tudjuk, hogy $e_{SI}(k) = (k+1) \cdot \frac{\gamma}{\tau} \cdot \frac{e_{k+1}u_{k+1}}{e_k u_k}$ és racionális törtfüggvények hányadosa is racionális törtfüggvény, ezért $e_{SI}(k)$ előáll racionális törtfüggvény alakban. Végül a $k = 0$ és $k = N$ esetekben,

könnyen látható, hogy $e_{SI}(0) = e_{SI}(N) = 0$, hiszen 0 vagy N beteg esetén nincsenek SI élek a gráfban. \square

Sajnos az u sajátvektor kiszámítása numerikusan lehetetlen nagy méretű gráfok esetén. Egy 10 csúcsú körgráfon azonban még kiszámíthatóak az $e_{SI}(k)$ értékek az eljárás alapján, és így az a_k ráták is. A kiszámolt együtthatókat felhasználva a (4.1) egyenletek megoldásait összevettük a (2.2) pontos alapegyenletekből nyert eredményekkel, ez látható a 4.3. ábrán. A kvázistacionárius állapotbeli betegszámot szinte teljesen pontosan megadja a (4.1) közelítő rendszer, ahol az u sajátvektor alapján végeztük a súlyozást. A becslés pontatlanságát a folyamat kezdeti időszakában az okozza, hogy az u sajátvektor - a folyamat későbbi szakaszában beálló - kvázi-egyensúlyi állapothoz tartozik. A módszer ezen gyengeségén igyekszünk javítani egy új eljárással a következő alszakaszban.



4.3. ábra. A betegek számának várható értéke körgráfon a (2.2) alapegyenletekből (piros görbe), a (4.5) súlyozással (fekete görbe) illetve a (4.6) súlyozással (cián görbe) a (4.1) redukált rendszerből és a (2.7) egyenletből (lila szaggatott görbe). $N = 10$, $\gamma = 1$, $\tau = 5$, 2 kezdeti beteggel indítva.

4.2.2. Becslés az SI élek átlagos számára lineáris közelítéssel

Az előző alszakaszban az $e_{SI}(k)$ értékek meghatározása a kvázi-sajátvektor alapján történt, így a módszer a járványterjedés kezdeti szakaszában nem ad jól használható eredményt. Most a

betegség terjedésének az elindulására fogunk koncentrálni és feltesszük, hogy a folyamat kezdeti feltétele egymáshoz közeli betegeket ad meg. Abból a megfigyelésből indulunk ki, hogy a fertőzött csúcsok egy többé-kevésbé összefüggő csoportot alkotnak a körgráf mentén, néhány egészséges csúccsal a csoport belsejében a visszagyógyulások következtében, azonban a már megfertőzött részgráfon egészséges szigetek nem alakulnak (újra) ki. Észrevehető, hogy adott hosszú, már fertőzött részén a gráfnak a betegek mennyisége állandó. Ezt úgy is fogalmazhatjuk, hogy az adott hosszú csoporton belül lévő spontán gyógyulások egyensúlyban vannak a belső megbetegedésekkel, azaz a (2.7) egyenletet alapul véve fennáll $\tau \bar{e}_{SI}(k) = \gamma k$, ahol $\bar{e}_{SI}(k)$ jelöli az SI élek átlagos számát a k beteget tartalmazó csoporton belül. Ebből az következik, hogy $e_{SI}(k)$ lineárisan függ k -tól és ennek a lineáris függvénynek a meredeksége γ/τ . Továbbá nyilvánvaló, hogy egy beteg esetén az SI élek száma 2, azaz $e_{SI}(1) = 2$, aminek segítségével már felírható az $e_{SI}(k)$ lineáris függvény:

$$e_{SI}(k) = 2 + \frac{\gamma}{\tau}(k-1) \quad \text{és} \quad a_k = \tau(2 + \frac{\gamma}{\tau}(k-1)). \quad (4.10)$$

A járványterjedés második szakaszában, amikor a betegek száma már majdnem eléri a folyamat kvázi-stacionárius állapotát, a fenti lineáris becslés már nem ad jó eredményt. Ekkor már a (4.5) becslés $n = 2$ mellett sokkal pontosabb közelítést nyújt, mivel ebben a szakaszban már úgy tekinthetjük, hogy a betegek és az egészségesek véletlenszerűen vannak elszórva a gráfon. Így legyen

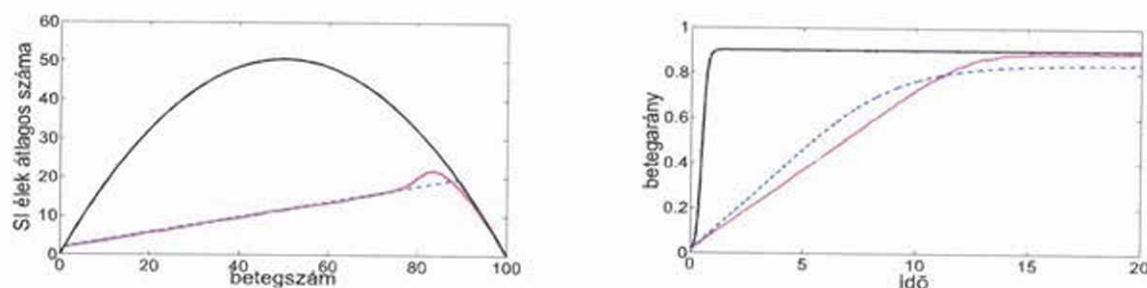
$$a_k = \tau \frac{2}{N-1} \cdot k \cdot (N-k), \quad (4.11)$$

ha $k > k_0$ és k_0 -t definiáljuk a (4.11) parabola és (4.10) lineáris egyenes metszéspontjaként, azaz k_0 legyen a

$$2 + \frac{\gamma}{\tau}(k_0 - 1) = \frac{2}{N-1} \cdot k_0 \cdot (N - k_0)$$

egyenlet megoldása és a (4.10) képlet csak $k \leq k_0$ esetén legyen érvényes.

A 4.4. ábrán ábrázoltuk az $e_{SI}(k)$ értékeket k függvényében a (4.10) és a (4.11) képletek alapján, illetve a szimulációból nyert $e_{SI}(k)$ értékekkel összevetettük ezt az elméleti görbét a baloldali ábrán. A jobboldali ábrán az átlagos betegszám látható a szimulációból kimérve illetve a (4.1) alapegyenletek alapján, felhasználva a (4.11) és a (4.10) képletekből kiszámított együtthatókat. Ennek a módszernek az előnye, hogy nagy méretű gráfon is használható az előző eljárással ellentétben, azonban a kvázi-egyensúlyi állapotban az átlagos betegszám értékét nem találja el pontosan.



4.4. ábra. Baloldali ábra: Az SI élek átlagos száma a betegszám függvényében körgráfon a szimulációból kimérve (piros görbe), a (4.10) és a (4.11) képletek alapján (kék szaggatott görbe) és a (4.5) képlet alapján (fekete görbe). Jobboldali ábra: A betegek átlagos aránya a teljes populációban körgráfon szintén a szimulációból kimérve (piros görbe), a (4.10) és a (4.11) képletek alapján (kék szaggatott görbe) és a (4.5) képlet alapján (fekete görbe). $N = 100$, $\gamma = 1$, $\tau = 5$. 500 szimuláció átlagából, 1 kezdeti beteggel.

4.3. Alapegyenletek a front hosszára

Ebben a fejezetben egy új eljárást adunk az átlagos betegszám kiszámítására, arra a megfigyelésre alapozva, hogy a betegség frontszerűen halad végig a körgráfon, ha a kezdeti feltétel egymáshoz közel elhelyezkedő beteg csúcsokból indítja el a folyamatot. Először úgy tekintünk a frontra, mint a körgráf egy összefüggő részére, amin csak betegek találhatók. A front két szélén következhet be változás: egyrészt egy új csúcs megfertőződésével a front növekedhet, másrészt a front szélén lévő betegek meggyógyulása által csökkenthet a hossza. A front belsejében végbemenő visszagyógyulásokat később vizsgáljuk majd meg. Bevezetünk egy folytonos paraméterű Markov-láncot a front hosszának a leírására, és az állapottérnek az $\{1, \dots, N\}$ halmazzt választjuk. Azt az esetet, amikor a front hossza 0, kizárjuk a vizsgálatból, ugyanis ez akkor fordulna elő, ha a kezdeti beteg csúcs még azelőtt meggyógyulna, mielőtt megfertőzné a szomszédait, de a paraméterek megfelelő beállításával - azaz ha a τ/γ arány megfelelően nagy - csak a betegség elterjedését vizsgáljuk a gráfon, ahogy már a 4.1. szakaszban megjegyeztük. Legyen $q_k(t)$, $k = 1, \dots, N$ annak a valószínűsége, hogy a front k hosszú a t időpontban, azaz pontosan k beteg csúcs helyezkedik el egymás mellett a gráfon. A rendszer a k -adik állapotból a $k + 1$ -edik állapotba juthat a front bármely szélén végbemenő fertőződés következtében 2τ rátával, illetve a $k - 1$ -edik állapotba gyógyuláskor 2γ rátával. A differenciálegyenlet-rendszer

a következő:

$$\begin{aligned}
 \dot{q}_1(t) &= -2\tau q_1(t) + 2\gamma q_2(t), \\
 &\vdots \\
 \dot{q}_k(t) &= 2\tau q_{k-1} - 2(\tau + \gamma)q_k(t) + 2\gamma q_{k+1}(t), \\
 &\vdots \\
 \dot{q}_N(t) &= 2\tau q_{N-1} - 2\gamma q_N(t).
 \end{aligned} \tag{4.12}$$

A rendszer megoldásából a front hosszának várható értéke az $m(t) = \sum_{k=1}^n k q_k(t)$ formulával számolható. A szimulációból vett megfigyelésekre hivatkozva kijelenthetjük, hogy a betegek számának várható értéke egyenes arányban áll a front hosszával, jelöljük az arányossági tényezőt α -val, azaz

$$[I](t) = \alpha m(t). \tag{4.13}$$

Ez az állítás azzal is magyarázható, hogy a fronton belüli betegeknek csak egy bizonyos hányada gyógyul vissza adott idő alatt, illetve egy bizonyos része fertőződik vissza, azaz a visszagyógyulás és a visszafertőződés egyensúlyban van a front belsejében. Ez azt jelenti, hogy a front belsejének bármely rögzített részén a betegszám állandó, tehát ott fennáll a (2.7) egyenlet alapján a $\tau[SI] = \gamma[I]$ összefüggés. A legegyszerűbb, gyakran használt becslés $[SI]$ -re $[I]$ függvényében a jól ismert (2.8) becslés egy változata: $[SI] = 2[I](m - [I])/m$ (lásd a 5.2.1. alszakaszt), ahol m a front hosszát jelöli, ami megegyezik a betegszámmal, hiszen feltettük, hogy a fronton csak betegek találhatók. Ezek alapján a két összefüggésből

$$[I](t) = \left(1 - \frac{\gamma}{2\tau}\right) m(t)$$

következik, vagyis $\alpha = 1 - \frac{\gamma}{2\tau}$ (összhangban az 5.2.1. alszakasz eredményeivel).

A 4.4. szakaszban található technikát felhasználva az α konstans ennél precízebben is meghatározható, méghozzá úgy, hogy $\alpha = I_s/m_s$, ahol I_s a betegek átlagos számának egyensúlyi állapota a (4.20) rendszer megoldása alapján és m_s a front hosszának egyensúlyi helyzete, amit alább fogunk meghatározni. Előrevetítve, például $\tau = 5$, $\gamma = 1$ esetén $1 - \frac{\gamma}{2\tau} = 0.9$, de $\alpha = I_s/m_s = 0.88$, tehát a $1 - \frac{\gamma}{2\tau}$ formula nem teljesen pontos. (Megjegyezzük, hogy I_s egyensúlyi értéket az előbbi ötlet mellett természetesen más közelítő rendszerekből, például a (2.9) egyenletekből is ki lehet számítani.)

A (4.12) differenciálegyenlet-rendszerből kiindulva levezethetünk egy analitikus képletet m_s -re, a front hosszának egyensúlyi értékére.

7. Állítás. A front hosszának várható értékét az $m(t) = \sum_{k=1}^n k q_k(t)$ formulával definiálva a (4.12) rendszer megoldásai alapján, a front hosszának egyensúlyi értéke:

$$m_s = N - \frac{\gamma}{\tau - \gamma} + \frac{N(\tau + \gamma)}{\left(\left(\frac{\tau}{\gamma}\right)^N - 1\right)(\tau - \gamma)}. \quad (4.14)$$

BIZONYÍTÁS. A (4.12) rendszer jobboldalát meghatározó együttható mátrixnak a 0 sajátértéke, hiszen az egyenletek összefüggőek, míg a többi sajátérték negatív valósrésztű. Tehát a folyamat stacionárius állapotát a 0-hoz tartozó $u = (u_1, \dots, u_N)^T$ sajátvektor adja, amit az alábbi lineáris egyenletrendszer megoldásaként kapunk:

$$\begin{pmatrix} -2\tau & 2\gamma & 0 & \dots & 0 \\ 2\tau & -(2\gamma + 2\tau) & 2\gamma & 0 & \dots & 0 \\ & & \vdots & & & \\ 0 & \dots & 0 & 2\tau & -(2\gamma + 2\tau) & 2\gamma \\ 0 & & \dots & 2\tau & -2\gamma \end{pmatrix} \cdot u = 0.$$

Összeadva az első k egyenletet, u koordinátáira fennáll $-2\tau u_k + 2\gamma u_{k+1} = 0$, vagyis $u_{k+1} = u_k \frac{\tau}{\gamma}$, $k = 1, \dots, N$. Azaz u koordinátái egy mértani sorozat egymást követő tagjai $\frac{\tau}{\gamma}$ kvócienssel. Mivel u koordinátái egy teljes eseményrendszer valószínűségei, ezért

$$1 = \sum_{k=1}^N u_k = u_1 \sum_{k=1}^N \left(\frac{\tau}{\gamma}\right)^{k-1} = u_1 \frac{\left(\frac{\tau}{\gamma}\right)^N - 1}{\frac{\tau}{\gamma} - 1}.$$

Emiatt az egyensúlyi állapot

$$u_1 = \frac{\frac{\tau}{\gamma} - 1}{\left(\frac{\tau}{\gamma}\right)^N - 1}, \quad u_k = u_1 \left(\frac{\tau}{\gamma}\right)^{k-1}. \quad (4.15)$$

Ezek alapján kiszámítható a fronthossz várható értéke:

$$m_s = \sum_{k=1}^N k u_k = u_1 \cdot \sum_{k=1}^N k \left(\frac{\tau}{\gamma}\right)^{k-1} = \frac{1 - (1 + N)\left(\frac{\tau}{\gamma}\right)^N + N\left(\frac{\tau}{\gamma}\right)^{N+1}}{\left(\left(\frac{\tau}{\gamma}\right)^N - 1\right)\left(\frac{\tau}{\gamma} - 1\right)},$$

ahol felhasználtuk, hogy

$$\sum_{k=1}^N k x^{k-1} = \frac{1 - (1 + N)x^N + Nx^{N+1}}{(x - 1)^2},$$

ami az $f(x) = \sum_{k=1}^N x^k = x \cdot \frac{x^N - 1}{x - 1}$ képlet kétoldali deriválásával jön ki. A fenti formulát átalakítva adódik, hogy

$$m_s = N - \frac{\gamma}{\tau - \gamma} + \frac{N(\tau + \gamma)}{\left(\left(\frac{\tau}{\gamma}\right)^N - 1\right)(\tau - \gamma)}.$$

□

A képlet jogosultságát a numerikus szimulációk igazolják. A 4.5. ábrán látható elméleti görbét a következő eljárással kaptuk. Először megoldottuk a (4.12) egyenleteket, majd kiszámítottuk a fronthossz várható értékét az $m(t) = \sum_{k=1}^n k q_k(t)$ képlet segítségével, végül az $[I]$ átlagos betegszámot a (4.13) összefüggésből nyertük $\alpha = I_s/m_s$ mellett, ahol I_s -t a (4.20) rendszer megoldásából, míg m_s -t a (4.14) képletből kaptuk.

A továbbiakban a célunk egyetlen közelítő differenciálegyenlet felírása a front hosszára, kihagyva az N egyenletet tartalmazó (4.12) rendszer megoldásainak kiszámítását. Az m függvényt lederiválva a

$$\dot{m}(t) = \sum_{k=1}^n k \dot{q}_k(t) = 2(\tau - \gamma) \sum_{k=1}^n q_k(t) + 2\tau q_1(t) - 2\gamma q_N(t) = 2(\tau - \gamma) + 2\gamma q_1(t) - 2\tau q_N(t). \quad (4.16)$$

differenciálegyenlethez jutunk. Ezt az egyenletet akkor lehet lezárni, ha a q_1 és a q_N függvényeket kifejezzük m segítségével, azaz találunk olyan L_1 és L_2 függvényeket, melyekre

$$q_1 = L_1(m), \quad q_N = L_N(m).$$

A szimulációs megfigyelésekre támaszkodva, az L_1 függvény értékei rövid front esetén még közel vannak az 1-hez, azonban ahogy a front átlagos hossza növekedésnek indul, a q_1 valószínűség 0-ig csökken. Ugyanígy, ahogy a front átlagos hosszát növeljük, az L_N függvény egy 0-ból induló, monoton növekvő függvény lesz, hiszen annak a valószínűsége, hogy a front hossza N , egyre nagyobb. Ebből a gondolatból kiindulva az L_1 és L_N függvényeket szakaszosan lineáris függvényekkel közelítjük, úgy hogy kielégítsék a

$$L_1(1) = 1, \quad L_1(h_1 N) = 0, \quad L_N(h_2 N) = 0, \quad L_N(N) = u_N$$

feltételeket, ahol h_1 és h_2 mesterséges paraméterek a szimulációból becsülve és u_N a (4.15)-beli egyensúlyi értéke q_N -nek. Tehát legyen

$$L_1(m) := \begin{cases} \frac{m-1}{1-h_1 N} + 1, & \text{ha } 1 \leq m < h_1 N, \\ 0, & \text{ha } h_1 N \leq m < N, \end{cases}$$

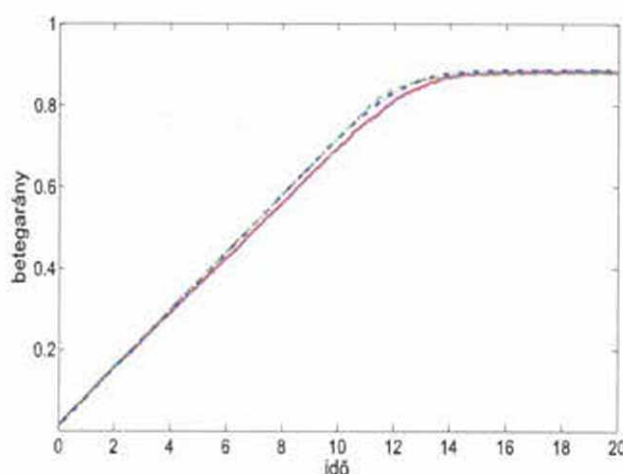
és

$$L_N(m) := \begin{cases} 0, & \text{ha } 1 < m < h_2 N, \\ \frac{u_N(m-h_2 N)}{N-h_2 N}, & \text{ha } h_2 N \leq m < N. \end{cases}$$

Így a (4.16) egyenletet a

$$\dot{m} = 2(\tau - \gamma) + 2\gamma L_1(m) - 2\tau L_N(m). \quad (4.17)$$

formába írhatjuk. Megint felhasználva a (4.13) összefüggést $\alpha = I_s/m_s$ mellett (I_s -t a (4.20) rendszer megoldásából, míg m_s -t a (4.14) képletből számítva), megkapjuk a betegek számának várható értékét. Ezt az elméleti görbét hasonlítjuk össze a szimulációval a 4.5. ábrán $h_1 = 0.1$ és $h_2 = 0.9$ paraméter választás esetén. Láthatóan ezen alszakasz módszerei működnek a legjobban körgráf esetén, ugyanis az egyezés kiváló. Még a kissé időigényesebb, első eljárás számítási igénye is elfogadható, hiszen a (4.12) rendszer N egyenletet tartalmaz, míg a (4.20) rendszer esetén $3N$ egyenlet megoldása szükséges I_s kiszámításához, azaz nagy méretű gráfon is jól használható.



4.5. ábra. A betegek átlagos aránya a teljes populációban körgráfon a szimulációból kimérve (piros görbe), a (4.12) rendszer és a (4.13) képlet alkalmazásával $\alpha = I_s/m_s$ mellett (kék szaggatott görbe), illetve a (4.17) egyenlet alapján $h_1 = 0.1$ és $h_2 = 0.9$ paraméterek mellett (zöld szaggatott görbe). $N = 100$, $\gamma = 1$, $\tau = 5$. 500 szimuláció átlagából, 1 kezdeti beteggel.

4.4. Az élek szintjén felírt egyenletek

Ebben az alszakaszban a Kieran Sharkey által kifejlesztett élek szintjén felírt egyenleteket [25] alkalmazzuk az átlagos betegszám meghatározására körgráf esetén. Az élek összes le-

hetséges állapotának valószínűségeire írunk fel differenciálegyenleteket. Ha a k -adik csúcs A állapotban van és a $k+1$ -edik csúcs B állapotban, akkor azt mondjuk, hogy a k -adik él (ami összeköti a k -adik és $k+1$ -edik csúcsot) AB állapotban van. Jelölje $y_{AB}^{k,k+1}(t)$, annak a valószínűségét, hogy a k -adik él az AB állapotban van a t időpontban, ahol $AB \in \{IS, SI, SS, II\}$. Nyilván

$$y_{IS}^{k,k+1} + y_{SI}^{k,k+1} + y_{SS}^{k,k+1} + y_{II}^{k,k+1} = 1 \quad (4.18)$$

fennáll $k = 1, \dots, N$ esetén, ahol az $(N, N+1)$ él legyen definíció szerint az $(N, 1)$ él. A [25] dolgozatban bevezetett modellt a körgráfra alkalmazva az élek állapotainak valószínűségeire az

$$\begin{aligned} \dot{y}_{SI}^{k,k+1} &= -(\gamma + \tau)y_{SI}^{k,k+1} + \gamma y_{II}^{k,k+1} + \tau y_{SSI}^{k,k+1,k+2} - \tau y_{ISI}^{k-1,k,k+1}, \\ \dot{y}_{IS}^{k,k+1} &= -(\gamma + \tau)y_{IS}^{k,k+1} + \gamma y_{II}^{k,k+1} + \tau y_{ISS}^{k-1,k,k+1} - \tau y_{IIS}^{k,k+1,k+2}, \\ \dot{y}_{SS}^{k,k+1} &= \gamma(y_{SI}^{k,k+1} + y_{IS}^{k,k+1}) - \tau y_{SSI}^{k,k+1,k+2} - \tau y_{ISS}^{k-1,k,k+1}, \\ \dot{y}_{II}^{k,k+1} &= -(\dot{y}_{SI}^{k,k+1} + \dot{y}_{IS}^{k,k+1} + \dot{y}_{SS}^{k,k+1}), \end{aligned} \quad (4.19)$$

differenciálegyenlet-rendszert kapjuk, ahol $y_{ABC}^{k-1,k,k+1}$ jelöli annak a valószínűségét, hogy a $k-1, k, k+1$ él az ABC állapotban van. Jegyezzük meg, hogy az utolsó egyenlet elhagyható a (4.18) összefüggőségi tulajdonság miatt.

A (4.19) differenciálegyenlet-rendszert a Kirkwood becslés [25] segítségével zárhatjuk le:

$$y_{ABC}^{k-1,k,k+1} = y_{AB}^{k-1,k} \frac{y_{BC}^{k,k+1}}{y_{BC}^{k,k+1} + y_{B-C}^{k,k+1}} = \frac{y_{AB}^{k-1,k}}{y_{AB}^{k-1,k} + y_{-AB}^{k-1,k}} y_{BC}^{k,k+1},$$

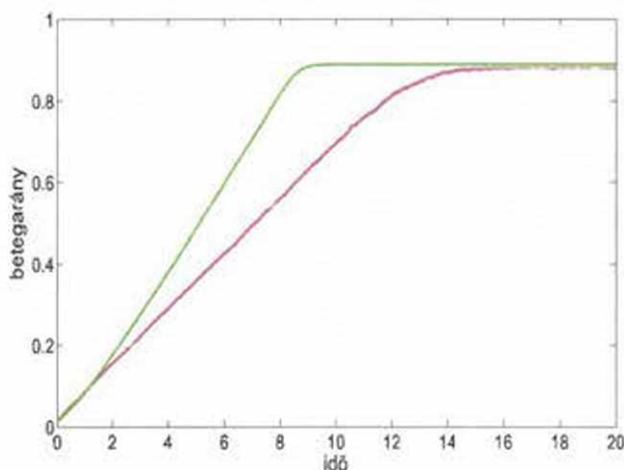
ahol például $y_{B-C}^{k,k+1}$ azt a valószínűséget jelöli, hogy a k -adik csúcs B állapotú, és a $k+1$ -edik csúcs nem a C állapotban van, a többi jelölés értelemszerűen következik ez alapján.

Behelyettesítve ezen becsléseket a (4.19) egyenletekbe, a következőt kapjuk:

$$\begin{aligned} \dot{y}_{SI}^{k,k+1} &= -(\gamma + \tau)y_{SI}^{k,k+1} + \gamma y_{II}^{k,k+1} + \tau y_{SS}^{k,k+1} \frac{y_{SI}^{k+1,k+2}}{y_{SI}^{k+1,k+2} + y_{SS}^{k+1,k+2}} - \tau \frac{y_{IS}^{k-1,k}}{y_{IS}^{k-1,k} + y_{SS}^{k-1,k}} y_{SI}^{k,k+1}, \\ \dot{y}_{IS}^{k,k+1} &= -(\gamma + \tau)y_{IS}^{k,k+1} + \gamma y_{II}^{k,k+1} + \tau \frac{y_{IS}^{k-1,k}}{y_{IS}^{k-1,k} + y_{SS}^{k-1,k}} y_{SS}^{k,k+1} - \tau y_{IS}^{k,k+1} \frac{y_{SI}^{k+1,k+2}}{y_{SI}^{k+1,k+2} + y_{SS}^{k+1,k+2}}, \\ \dot{y}_{SS}^{k,k+1} &= \gamma(y_{SI}^{k,k+1} + y_{IS}^{k,k+1}) - \tau y_{SS}^{k,k+1} \frac{y_{SI}^{k+1,k+2}}{y_{SI}^{k+1,k+2} + y_{SS}^{k+1,k+2}} - \tau \frac{y_{IS}^{k-1,k}}{y_{IS}^{k-1,k} + y_{SS}^{k-1,k}} y_{SS}^{k,k+1}. \end{aligned} \quad (4.20)$$

A (4.20) rendszer numerikusan megoldható, ugyanis az egyenletek száma N nagyságrendű. A $y_{II}^{k,k+1}$ valószínűségeket a $y_{II}^{k,k+1} = 1 - y_{SI}^{k,k+1} - y_{IS}^{k,k+1} - y_{SS}^{k,k+1}$ képletből számítottuk ki. Annak a valószínűségét, hogy a k -adik csúcs beteg, y_I^k -vel jelöltük és a (4.20) rendszer

megoldásaiból a $y_I^k = \frac{1}{2}(y_{SI}^{k-1,k} + y_{IS}^{k,k+1} + y_{II}^{k-1,k} + y_{II}^{k,k+1})$ formula segítségével számoltuk ki. Az átlagos betegszámot az $[I] = \sum_{k=1}^n y_I^k$ képlet adta meg. A 4.6. ábrán összevetettük a módszerből kapott betegszámarányt a szimuláció eredményével. Látható, hogy ez az eljárás felülbecsüli a betegség terjedési sebességét, de jól meghatározza a betegek arányát a teljes populációban. Emiatt volt érdemes ezt az egyensúlyi értéket felhasználni α meghatározásához a 4.3. szakaszban.



4.6. ábra. A betegek arányának várható értéke az idő függvényében körgráfon a szimulációból (piros görbe) és a (4.20) rendszerből (zöld görbe). $N = 100$, $\gamma = 1$, $\tau = 5$. 500 szimuláció átlagából, 1 kezdeti beteggel.

4.5. Összefoglalás

A fejezet eredménye alacsony dimenziós közelítő differenciálegyenlet-rendszerek előállítása, melyek megőrzik a főbb tulajdonságait az SIS típusú járványterjedésnek körgráf esetén. A kutatást a gráf speciális struktúrája motiválta, kiemelve azt a tényt, hogy a gyakran alkalmazott (2.7) és a (2.9) egyenletek, kiugróan rosszul működnek körgráf esetén. Három különböző eljárást kínáltunk a betegek átlagos számának előállítására az idő függvényében.

Az első módszerrel a betegek száma alapján írtunk fel $N + 1$ differenciálegyenletet. A átmeneti ráták meghatározása során különböző becsléseket adtunk az $e_{SI}(k)$ értékekre, azaz az SI élek átlagos számára k beteg esetén. Kétféle közelítő módszert mutattunk az SI élek várható értékének kiszámítására. Az első alszakaszban a 2^N dimenziós rendszer kvázi-stacionárius

állapotához tartozó sajátvektor alapján súlyozzuk a különböző állapotokban megjelenő SI értékeket. Sajnos nagy méretű gráfon a sajátvektor kiszámítása nem lehetséges, tekintve a rendszer méretét, de kis csúcsszámú gráfon a módszer jó közelítést ad. Ebben az alszakaszban tárgyalt állítások inkább elméleti jelentőségűek, gyakorlati hasznuk nagy méretű körgráfon csekély. A másik alszakaszban heurisztikus érveléssel egy egyszerű formulát adtunk meg az $e_{SI}(k)$ értékekre. Ezt a képletet bármilyen méretű körgráfon fel lehet írni, azonban pontatlan közelítést kapunk.

A második eljárás azon a megfigyelésen alapszik, hogy a betegség a körgráfon frontszerűen halad végig. Bevezettünk egy új Markov-láncot a front hosszának leírására. A Markov-folyamat állapottere csupán N elemű, ami egy N dimenziós rendszerhez vezet, így nagy csúcsszámú gráf esetén is könnyedén megoldható numerikusan. Nagyon precíz közelítést kapunk ebben az esetben, azonban megjegyezzük, hogy itt az átlagos betegszámot a front hosszának konstansszorosaként kapjuk meg. Az eljárás pontossága ettől a konstanstól függ, azonban ezt nehéz analitikusan meghatározni.

Az utolsó technikaként egy ismert eljárást alkalmazunk körgráf esetére. Ez a módszer nagy N -re is működik, tekintve hogy az egyenletek száma N nagyságrendű. A szimulációval való összevetésből arra következtetünk, hogy a rendszer megoldása az egyensúlyi helyzetet meglehetősen pontosan eltalálja, azonban a járványterjedés kezdeti szakaszában rossz közelítést ad.

A különböző eljárások egyfajta összehasonlítására kiszámoltuk a betegszám egyensúlyi állapotát ötféleképpen: a szimulációból, a (2.7) és a (2.9) egyenletekből, a (4.1) redukált egyenletekből a (4.10) és a (4.11) képletekből nyert rátákkal illetve a (4.20) rendszerből. Az 4.1. táblázatban rögzítettük az egyensúlyi állapotban a betegek átlagos számának arányát a teljes populációhoz képest. Megfigyelhető, hogy a (2.7) és a (4.1) redukált alapegyenletek hasonló eredményre jutnak, hiszen ezek az egyenletek a csúcsok szintjén épülnek fel és a lezárások a párokra vonatkozó közelítésekkel élnek. Ugyanígy egymáshoz közeli eredményt adnak a (2.9) és a (4.20) rendszerek, ami megint azzal magyarázható, hogy ezekben az egyenletekben pedig mindkét esetben az élek állapotait tekintjük a változóknak.

4.1. táblázat. Az átlagos betegarány az egyensúlyi állapotban különböző τ értékek mellett a szimulációból, a (2.7) és a (2.9) egyenletekből, a (4.1) redukált egyenletekből a (4.10) és a (4.11) képletekből nyert rátákkal illetve a (4.20) rendszerből.

τ	Fertőzöttek aránya az egyensúlyi állapotban				
	szimuláció	(2.7)	(2.9)	(4.1)	(4.20)
2	0.6127	0.7512	0.6666	0.7496	0.6667
5	0.8763	0.9005	0.8889	0.8999	0.8889
10	0.9445	0.9503	0.9474	0.95	0.9474

5. fejezet

Alacsony dimenziós közelítő rendszerek analízise

5.1. Motiváció

Ebben a fejezetben a (2.2) pontos alapegyenlet-rendszer kiváltására alkalmas alacsony dimenziós közelítő rendszerekkel fogunk foglalkozni, azonban új eljárások kidolgozása helyett, már ismert approximativ differenciálegyenlet-rendszerek analitikus vizsgálatát szeretnénk bemutatni, illetve elvégezni. A dolgozat bevezetőjében említést tettünk alacsony dimenziós közelítő modellekről, mint a (2.7) és a (2.9) egyenletek. Először néhány ismert és fontos állítást foglalkozunk meg ezen rendszerekről, illetve ezek - heterogén foksámeloszlású gráf esetén alkalmazott - változatairól. Az említett rendszerekről bővebben a [16, 14, 23] munkákból tájékozódhatunk. Ezek után egy újabban kidolgozott rendszer, a párok szintjén felírt kompakt egyenletek viselkedését elemezzük [16].

A továbbiakban tekintsünk egy N csúcsú gráfot SIS dinamika mellett. A 2.4. szakaszban leírt függvényesaládot alkalmazzuk az átmenetek meghatározására, azaz bármely fertőzött csúcs γ rátával gyógyul meg, illetve egy egészséges csúcs τl rátával betegszik meg l fertőzött szomszéd esetén, ahol $\gamma > 0$, $\tau > 0$ paraméterek.

5.2. Átlagoláson alapuló közelítő modellek

Ezek, a bevezetőben említett, alacsony dimenziós közelítő egyenletek nagyon egyszerű alakúak, ezért könnyen elvégezhető az analitikus vizsgálatuk, azonban a legtöbb gráf típus esetén nem adnak jó közelítést.

5.2.1. Az átlagoláson alapuló egyenlet

Amennyiben csak a betegek átlagos számára vagyunk kíváncsiak egy N csúcsú gráfon, akkor a 2^N méretű (2.2) alapegyenlet-rendszer helyett elegendő lehet az átlagolással kapott (2.7) egyenlet(ek) vizsgálata. Az ilyen típusú egyenletek az egyes változók átlagolt értékeivel dolgoznak, mint az egészséges csúcsok átlagos száma: $[S]$, a beteg csúcsok átlagos száma: $[I]$, illetve az SI élek átlagos száma: $[SI]$, azaz a bevezetőben ismertetett jelölések mellett legyen

$$[S] = \sum_{k=0}^N (N-k) \sum_{i=1}^{d_k} x_{S_i^k}, \quad [I] = \sum_{k=0}^N k \sum_{i=1}^{d_k} x_{S_i^k} \quad \text{és} \quad [SI] = \sum_{k=0}^N \sum_{i=1}^{d_k} N_{SI}(S_i^k) x_{S_i^k}.$$

Ezen változókra fennáll a - már bevezetőben is említett -

$$\begin{aligned} \dot{[S]} &= \gamma[I] - \tau[SI], \\ \dot{[I]} &= \tau[SI] - \gamma[I] \end{aligned} \quad (5.1)$$

egyenletrendszer, ami levezethető az (2.2) pontos 2^N dimenziós differenciálegyenlet-rendszerből [26], azaz a (5.1) rendszer pontos összefüggést ad meg a betegek, illetve az egészségesek és az SI élek átlagos száma között. Természetesen a két egyenlet összefüggő, felhasználva az $[S] + [I] = N$ állítást, hiszen SIS dinamika esetén bármely csúcs csak S vagy I típusú lehet.

Az egyenlet lezárásához az $[SI]$ változóra felírunk egy becslést $[I]$ függvényében. Ez leggyakrabban a

$$[SI] \approx \frac{n}{N} [S][I] = \frac{n}{N} (N - [I])[I] \quad (5.2)$$

becslés, ugyanis a betegek aránya a gráfban $\frac{[I]}{N}$, így egy egészséges csúcsból átlagosan $n \frac{[I]}{N}$ él vezet beteg csúcshoz, ha az átlagos fokszám n , így az összes egészséges csúcshoz átlagosan $\frac{n}{N} [S][I]$ darab SI él tartozik. Ennél kicsit pontosabb a

$$[SI] \approx \frac{n}{N-1} [S][I] = \frac{n}{N-1} (N - [I])[I] \quad (5.3)$$

közelítés, ahol egy egészséges csúcs kiválasztása után, a betegek aránya a gráfban $\frac{[I]}{N-1}$, így egy egészséges csúcsból átlagosan $n \frac{[I]}{N-1}$ él vezet beteghez.

Felhasználva az (5.1) egyenletek összefüggését és - az egyszerűség kedvéért - az (5.2) lezárási formulát az egyenletrendszer helyett a

$$\dot{[I]} = \tau \frac{n}{N} (N - [I])[I] - \gamma [I] \quad (5.4)$$

közelítő, átlagoláson alapuló egyenletet (homogeneous mean-field) tekintjük, ahol a rövidség kedvéért az eredeti jelöléseket használjuk a lezárt rendszer közelítő megoldásaira is. Megjegyezzük, hogy a fejezet folyamán bemutatásra kerülő további lezárt egyenletrendszerek esetén is így járunk el a jelölések tekintetében.

Ezen egy dimenziós rendszer viselkedésének a vizsgálata kifejezetten egyszerű, a legfontosabb megállapításokat foglaljuk össze a továbbiakban. Mivel az egyenletet járványterjedés modellezésére használjuk, így csak azon megoldások elemzésével fogunk foglalkozni, amikor a betegek száma nemnegatív.

8. Állítás. Legyen $\tau_c^{HomMF} = \frac{\gamma}{n}$. Az (5.4) egyenletnek van egy triviális egyensúlyi pontja: $[I] = 0$, azaz ilyenkor a gráf összes csúcsa egészséges. Emellett lehet egy endemikus egyensúlyi pontja is: $[I] = N(1 - \frac{\gamma}{n\tau})$, amennyiben ez az érték pozitív, azaz $\tau_c^{HomMF} < \tau$. Ha $\tau < \tau_c^{HomMF}$, akkor a triviális egyensúlyi pont stabil, míg ha $\tau_c^{HomMF} < \tau$, akkor a triviális egyensúlyi pont instabillá válik és a megjelenő endemikus egyensúlyi pont pedig stabil lesz. Tehát $\tau = \tau_c^{HomMF}$ esetén transzkritikus bifurkáció jelenik meg.

Természetesen, általános vizsgálat esetén, ha $\tau \neq \tau_c^{HomMF}$, akkor két egyensúlyi pont van, csak az egyik negatív, ha $\tau < \tau_c^{HomMF}$.

A következő részben az (5.4) rendszernél pontosabb párok szintjén felírt egyenletek rövid analízise található.

5.2.2. A párok szintjén felírt egyenletek

Most a bevezetőben ismertetett párok szintjén felírt (2.9) egyenletek elemzésével folytatjuk a sort, melyek szintén pontosan levezethetők a (2.2) alapegyenletekből [26]. A párok szintjén felírt egyenletek pontosabb modellt adhatnak az átlagolással kapott (5.4) egyenleteknél, ugyanis itt több, a rendszer működését leíró változóról kapunk információt.

Tekintsük a (2.9) rendszert a (2.10) lezárási formulák alkalmazása mellett:

$$\begin{aligned}
 \dot{[S]} &= \gamma[I] - \tau[SI], \\
 \dot{[I]} &= \tau[SI] - \gamma[I], \\
 \dot{[SI]} &= \gamma([II] - [SI]) + \tau \left(\frac{n-1}{n} \frac{[SS][SI]}{[S]} - \frac{n-1}{n} \frac{[SI]^2}{[S]} - [SI] \right), \\
 \dot{[SS]} &= 2\gamma[SI] + 2\tau \frac{n-1}{n} \frac{[SS][SI]}{[S]}, \\
 \dot{[II]} &= -2\gamma[II] + 2\tau \left(\frac{n-1}{n} \frac{[SI]^2}{[S]} + [SI] \right).
 \end{aligned} \tag{5.5}$$

Ezen párok szintjén felírt differenciálegyenlet-rendszer (homogeneous pairwise) analitikus vizsgálatával foglalkozunk a továbbiakban. Az élek összegére ismert $[SS] + 2[SI] + [II] = nN$ összefüggés mellett, ahol n jelöli az átlagos foksámot, felírhatóak az

$$n[S] = [SS] + [SI] \quad \text{és} \quad n[I] = [II] + [SI]$$

összefüggések is, hiszen például $n[S]$ megadja az egészséges csúcsokból kiinduló éleknek az átlagos számát, ami megegyezik az SS és az SI típusú élek átlagos számának összegével. Felhasználva az (5.5) rendszer előbb említett összefüggőségi tulajdonságait és az $[S] + [I] = N$ állítást, az (5.5) egyenlet az

$$\begin{aligned}
 \dot{[S]} &= \gamma N - (\gamma + n\tau)[S] + \tau[SS], \\
 \dot{[SS]} &= 2(n[S] - [SS]) \left(\gamma - \tau(n-1) \frac{[SS]}{n[S]} \right).
 \end{aligned} \tag{5.6}$$

kétdimenziós egyenletrendszerre redukálható, aminek analitikus vizsgálata már könnyedén elvégezhető.

9. Állítás. Legyen $\tau_c^{HomPW} = \frac{\gamma}{n-1}$. A (5.6) rendszer triviális egyensúlyi helyzete: $[S] = N$, $[SS] = nN$, ami stabil, ha $\tau < \tau_c^{HomPW}$. Amennyiben $\tau_c^{HomPW} < \tau$, akkor a triviális egyensúlyi állapot instabillá válik és megjelenik egy endemikus stabil egyensúlyi pont is: $[S] = N \frac{\gamma(n-1)}{\tau n(n-1) - \gamma}$, $[SS] = \frac{N}{\tau} \frac{\gamma^2 n}{\tau n(n-1) - \gamma}$. Ezen észrevételek alapján az (5.5) rendszerre is jellemzést adhatunk, pontosabban ha $\tau < \tau_c^{HomPW}$, akkor csak a csupa egészségesekből álló állapot lesz egyensúlyi helyzet, ha viszont $\tau_c^{HomPW} < \tau$, akkor megjelenik egy endemikus egyensúlyi állapot is és az stabil.

Reguláris véletlen gráf esetén, ha az n átlagos foksám magas, akkor az (5.4) és az (5.5) egyenletek jó közelítést adnak, ahogy a 4. fejezetben már említettük, azonban ha a gráf fokszámeloszlása heterogén, akkor a lezárási formulákban megjelenő átlagos foksám használata

nem ad elég információt a gráfstruktúráról, így ezen rendszerek eredménye pontatlan lesz. Heterogén foksámeloszlású gráf esetén érdemesebb a következő fejezetben ismertetett általánosabb modellekkel dolgozni a pontosabb közelítés érdekében.

5.3. Közelítő modellek heterogén foksámeloszlású hálózaton

A fejezetben heterogén foksámeloszlású gráfon fogunk dolgozni. Legyen a gráf N csúcsú, L a különböző fokok száma és jelölje N_l az n_l fokú csúcsok számát, vagyis a gráf foksámeloszlása: $p(n_l) = N_l/N$, $l = 1, 2, \dots, L$, és nyilván $N_1 + N_2 + \dots + N_L = N$.

5.3.1. Fokszám szerinti átlagoláson alapuló és párok szintjén felírt egyenletek

Heterogén foksámeloszlású gráfon, azaz ahol különböző fokú csúcsok is előfordulnak, az (5.1) egyenlet helyett érdemes a fokszám szerinti átlagoláson alapuló egyenleteket használni. Ekkor különböző fokú egészséges illetve beteg csúcsokra külön-külön írunk fel differenciálegyenleteket. Jelölje $[S_l]$ az n_l fokú egészséges csúcsok átlagos számát, és ugyanígy $[I_l]$ az n_l fokú fertőzött csúcsok átlagos számát, $l = 1, 2, \dots, L$. Ekkor a fokszám szerinti átlagoláson alapuló egyenletrendszer alakja a következő:

$$\begin{aligned} \dot{[S_l]} &= \gamma[I_l] - \tau[S_l I], \quad l = 1, \dots, L, \\ \dot{[I_l]} &= \tau[S_l I] - \gamma[I_l], \quad l = 1, \dots, L, \end{aligned} \quad (5.7)$$

ahol értelemszerűen $[S_l I]$ -vel jelöljük az n_l fokú egészséges csúcsokból tetszőleges fokú beteg csúcsba kiinduló élek átlagos számát. Szemléletesen az egyenletek azt fejezik ki, hogy - például az első egyenletek esetén - az n_l fokú beteg csúcsok átlagos számával arányosan nő az n_l fokú egészséges csúcsok átlagos száma, γ rátával gyógyulás által, míg az $S_l I$ típusú élek átlagos számával arányosan csökken, τ rátával. Mivel $[S_l] + [I_l] = N_l$ itt is fennáll, ezért felhasználva az egyenletek összefüggőségét, az egészséges csúcsokra felírt egyenletek elhagyhatóak.

Az (5.7) egyenletek esetén is szükséges a rendszer lezárása, így az $[S_l I]$ változókra felírunk egy közelítő képletet. Az összes csúcsból kiinduló fél-élek száma $\sum_{l=1}^L n_l N_l$, míg a beteg

csúcsokból kiinduló fél-élek száma $\sum_{l=1}^L n_l[I_l]$. Ezek alapján annak a valószínűsége, hogy egy tetszőleges fél-él beteg csúcsból indul ki:

$$Q_{MF} = \frac{\sum_{l=1}^L n_l[I_l]}{\sum_{l=1}^L n_l N_l}.$$

Így az olyan n_l fokú egészséges csúcsból kiinduló élek átlagos száma, melyek beteg csúcshoz csatlakoznak:

$$[S_l I] \approx n_l [S_l] Q_{MF}.$$

Ezt a becslést használjuk fel a rendszer lezárására, így a lezárt, foksám szerinti átlagoláson alapuló egyenletrendszer (heterogeneous mean-field) alakja:

$$[\dot{I}_l] = \tau n_l (N_l - [I_l]) Q_{MF} - \gamma [I_l], \quad l = 1, \dots, L. \quad (5.8)$$

Vezessük be a következő jelöléseket a foksám eloszlás várható értékére és második momentumára:

$$\langle n \rangle = \frac{1}{N} \sum_{l=1}^L l N_l \text{ és } \langle n^2 \rangle = \frac{1}{N} \sum_{l=1}^L l^2 N_l.$$

Ezen jelölések mellett az (5.8) egyenletekre igaz a következő állítás.

10. Állítás. Tekintsük az (5.8) rendszert és legyen $\tau_c^{HetMF} = \gamma \frac{\langle n \rangle}{\langle n^2 \rangle}$. A rendszer triviális egyensúlyi állapota: $[I_l] = 0$, $l = 1, \dots, L$. Ez stabil, ha $\tau < \tau_c^{HetMF}$, illetve $\tau_c^{HetMF} < \tau$ esetén instabillá válik, és megjelenik egy endemikus egyensúlyi pont, ami stabil.

A foksám szerinti párok szintjén felírt egyenleteknek (heterogeneous pairwise) nem szentelünk külön alszakaszt, itt csak ismertetjük az egyenleteket, ugyanis ezen bonyolult rendszer analízise helyett egy egyszerűbb modell elemzésével szeretnénk foglalkozni a következő részben. Ezen egyenletek esetén az (5.1) rendszert bővítjük a különböző foksámmal rendelkező párokra felírt egyenletekkel:

$$\begin{aligned} [\dot{S}_l] &= \gamma [I_l] - \tau [S_l I], \\ [\dot{I}_l] &= \tau [S_l I] - \gamma [I_l], \\ [\dot{S}_k I_l] &= \gamma ([I_k I_l] - [S_k I_l]) + \tau ([S_k S_l I] - [I S_k I_l] - [S_k I_l]), \\ [\dot{S}_k S_l] &= \gamma ([S_k I_l] + [I_k S_l]) - \tau ([S_k S_l I] + [I S_k S_l]), \\ [\dot{I}_k I_l] &= \tau ([S_k I_l] + [I_k S_l]) - 2\gamma [I_k I_l] + \tau ([I S_k I_l] + [I_k S_l I]), \end{aligned} \quad (5.9)$$

ahol értelemszerűen $[A_k B_l]$ jelöli az A típusú n_k fokú és B típusú n_l fokú csúcsokat összekötő élek átlagos számát és az

$$[A_k S_l I] \approx \frac{n_l - 1}{n_l} \frac{[A_k S_l][S_l I]}{[S_l]} \quad \text{és} \quad [I S_k A_l] \approx \frac{n_k - 1}{n_k} \frac{[I S_k][S_k A_l]}{[S_k]} \quad (5.10)$$

közelítésekkel élünk a hármasokra a rendszer lezárásaképpen, $A, B \in \{S, I\}$.

Ez a rendszer $3L + 2L^2$ egyenletet tartalmaz, így analitikus vizsgálata nehézkes. Ezért egy másik lezárási formulát alkalmazunk az (5.9) egyenletrendszerre, aminek segítségével a párokra felírt $L^2 + L(L + 1)$ darab egyenlet típusonként összeadható, így egy kisebb dimenziós rendszerhez jutunk.

5.3.2. Párok szintjén felírt kompakt egyenletrendszer

Az (5.9) egyenletek redukálására felhasználunk egy új típusú lezárási formulát. Ehhez az

$$[S_l I] \approx [S I] \frac{n_l [S_l]}{\sum_{k=1}^L n_k [S_k]} \quad (5.11)$$

közelítést használjuk, ugyanis az n_l fokú egészségesekből kiinduló fél-élek számát elosztva az összes egészségesből kiinduló fél-éllel, megkapjuk az n_l fokú egészséges fél-élek arányát az összes egészséges fél-élhez képest. Ezt az arányt megszorozva az $[S I]$ élek átlagos számával egy becslést kapunk $[S_l I]$ -re. Ezt a formulát és az (5.10) közelítést alkalmazva a hármasok lezárására

$$[A_k S_l I] \approx (n_l - 1) [A_k S_l] \frac{[S I]}{S_s} \quad (5.12)$$

adódik, ahol $S_s = \sum_{k=1}^L n_k [S_k]$.

Felhasználva a $\sum_{k=1}^L [A_k S_l] = [A S_l]$ összefüggést és az (5.12) becslést

$$\sum_{k=1}^L [A_k S_l I] \approx (n_l - 1) \sum_{k=1}^L [A_k S_l] \frac{[S I]}{S_s} = (n_l - 1) [A S_l] \frac{[S I]}{S_s} \approx (n_l - 1) n_l [S_l] \frac{[A S][S I]}{S_s^2}$$

ahol kihasználtuk az $[A S_l] \approx [A S] \frac{n_l [S_l]}{\sum_{k=1}^L n_k [S_k]}$ közelítést, ami hasonlóan indokolható, mint az (5.11) összefüggés. Így

$$\sum_{l=1}^L \sum_{k=1}^L [A_l S_k I] \approx \frac{[A S][S I]}{S_s^2} \sum_{l=1}^L (n_l - 1) n_l [S_l]. \quad (5.13)$$

A (5.9) rendszer párokra felírt egyenleteit típusonként összeadva $k = 1$ -től L -ig és $l = 1$ -től L -ig, illetve alkalmazva az (5.11) és az (5.13) becsléseket és hogy $\sum_{l=1}^L \sum_{k=1}^L [A_k B_l] = [A B]$

megkapjuk a párok szintjén felírt kompakt egyenleteket (compact pairwise):

$$[\dot{S}_l] = \gamma[I_l] - \tau n_l[S_l] \frac{[SI]}{S_s}, \quad l = 1, \dots, L \quad (5.14)$$

$$[\dot{I}_l] = \tau n_l[S_l] \frac{[SI]}{S_s} - \gamma[I_l], \quad l = 1, \dots, L \quad (5.15)$$

$$[\dot{S}] = \gamma([II] - [SI]) + \tau([SS] - [SI])[SI]Q_{CP} - \tau[SI], \quad (5.16)$$

$$[\dot{S}] = 2\gamma[SI] - 2\tau[SS][SI]Q_{CP}, \quad (5.17)$$

$$[\dot{II}] = 2\tau[SI] - 2\gamma[II] + 2\tau[SI]^2Q_{CP}, \quad (5.18)$$

ahol

$$S_s = \sum_{l=1}^L n_l[S_l] = [SS] + [SI], \quad Q_{CP} = \frac{1}{S_s^2} \sum_{l=1}^L (n_l - 1)n_l[S_l].$$

Megjegyezzük, hogy az eddigi hagyományoknak megfelelően az eredeti jelöléseket alkalmazzuk az egyenletek közelítő megoldásaira is.

Az $[S_l] + [I_l] = N_l$ összefüggés miatt az (5.14) és az (5.15) egyenletek helyett az

$$[\dot{S}_l] = \gamma(N_l - [S_l]) - \tau n_l[S_l] \frac{[SI]}{S_s}, \quad l = 1, \dots, L \quad (5.19)$$

egyenleteket fogjuk használni az (5.16)-(5.18) egyenletek mellett. A csökkentett méretű rendszer csak $L + 3$ egyenletet tartalmaz, így könnyebben kezelhető az (5.9) rendszernél. A következőkben az (5.16)-(5.19) modell analitikus vizsgálatával foglalkozunk.

11. Állítás. Az (5.16)-(5.19) rendszer triviális egyensúlyi állapota: $[S_l] = N_l$, $l = 1, \dots, L$, $[SI] = 0$, $[SS] = \langle n \rangle \cdot N$, $[II] = 0$. Legyen $\tau_c^{CPW} = \gamma \frac{\langle n \rangle}{\langle n^2 \rangle - \langle n \rangle}$. Ekkor a triviális egyensúlyi pont (aszimptotikusan) stabil $\tau < \tau_c^{CPW}$ esetén, illetve instabil, ha $\tau_c^{CPW} < \tau$.

BIZONYÍTÁS. Egyszerű számolással látható, hogy a fenti értékek egyensúlyi állapotot adnak. A stabilitást linearizálással fogjuk belátni, amihez az egyenlet jobboldalát kell deriválni minden változó szerint, majd beírni a Jacobi-mátrixba az egyensúlyi értékeket. Az (5.19) egyenletet deriválva $[S_k]$ szerint, majd behelyettesítve az egyensúlyi értékeket, az eredmény nulla, ha $k \neq l$. Ha viszont $[S_l]$ szerint deriváljuk az l -edik egyenletet, akkor behelyettesítés után $-\gamma$ -t kapunk. Az (5.16), az (5.17) és az (5.18) egyenletek jobboldalai $[S_l]$ szerint deriválva, majd behelyettesítés után, szintén nullát adnak, $l = 1, \dots, L$. Viszont ha ugyanazeket az egyenleteket deriváljuk az $[SI]$, $[SS]$ és $[II]$ változók szerint, majd behelyettesítjük az egyensúlyi állapotot, akkor a

következő mátrixhoz jutunk:

$$P = \begin{pmatrix} \alpha & 0 & \gamma \\ \beta & 0 & 0 \\ 2\tau & 0 & -2\gamma \end{pmatrix},$$

ahol

$$\alpha = \tau \frac{\langle n^2 \rangle - \langle n \rangle}{\langle n \rangle} - (\tau + \gamma) \quad \text{és} \quad \beta = 2\gamma - 2\tau \frac{\langle n^2 \rangle - \langle n \rangle}{\langle n \rangle}.$$

Így a teljes rendszer Jacobi-mátrixa a triviális egyensúlyi pontban:

$$J = \begin{pmatrix} -\gamma I & R \\ 0 & P \end{pmatrix},$$

ahol R a megfelelő $L \times 3$ -as mátrix, 0 jelöli a $3 \times L$ -es nullmátrixot és I az $L \times L$ -es egységmátrix. A J mátrixnak a $-\gamma$ L -szeres sajátértéke, ezenkívül sajátérték még a 0 és a

$$\lambda^2 + \lambda(2\gamma - \alpha) - 2\gamma(\alpha + \tau)$$

másodfokú polinom gyökei. A nulla sajátérték csak az (5.16), az (5.17) és az (5.18) egyenletek összefüggőségének köszönhető, - ugyanis belátható, hogy $[SS] + 2[SI] + [II] = nN$, - így ez a sajátérték nem befolyásolja a stabilitást. A kvadratikus polinom sajátértékeinek valós része pontosan akkor negatív, ha $\alpha + \tau < 0$. Tehát a triviális egyensúlyi pont pontosan akkor (aszimptotikusan) stabil, ha $\tau < \tau_c^{CPW}$, ellenkező esetben pedig ha $\tau_c^{CPW} < \tau$, akkor instabillá válik.

□

12. Tétel. Ha $\tau_c^{CPW} < \tau$, ahol $\tau_c^{CPW} = \gamma \frac{\langle n \rangle}{\langle n^2 \rangle - \langle n \rangle}$, akkor a (5.16)-(5.19) kompakt pairwise modelnek létezik endemikus egyensúlyi állapota (azaz olyan, melyre $[S_l] < N_l$) és ez az egyensúlyi állapot egyértelmű.

BIZONYÍTÁS. Legyenek az $[S_l]$, az $[SI]$, az $[SS]$ és az $[II]$ változók értékei az egyensúlyi állapotban rendre $X_l, Z, U, V, l = 1, \dots, L$ és az egyszerűség kedvéért az átlagos foksámot n -el fogjuk jelölni helyenként, azaz $n := \langle n \rangle$.

Az (5.19) egyenletekbe behelyettesítve az egyensúlyi értékeket:

$$\gamma N_l = X_l \left(\gamma + \tau n_l \frac{Z}{Z + U} \right), \quad l = 1, \dots, L.$$

Kifejezve az X_l tagokat, beszorozva az egyenleteket $n_l(n_l - 1)$ -el, majd összeadva az egyenleteket

$$\gamma \sum_{l=1}^L \frac{n_l(n_l - 1)N_l}{\gamma + \tau n_l \frac{Z}{Z+U}} = \sum_{l=1}^L n_l(n_l - 1)X_l = Q_{CP}(U + Z)^2.$$

Felhasználva az egyensúlyi helyzetből következő $\tau Q_{CP}U = \gamma$ összefüggést az (5.17) egyenletből és elosztva $Q_{CP}(U + Z)^2$ -vel

$$1 = \frac{\tau U}{U + Z} \sum_{l=1}^L \frac{n_l(n_l - 1)N_l}{\gamma(U + Z) + \tau n_l Z}.$$

Az (5.18) egyenletből következik, hogy az egyensúlyi állapotban $\gamma V - \tau Z = \tau Z^2 Q_{CP}$ teljesül. Ezt és a $\tau Q_{CP}U = \gamma$ összefüggést összekombinálva az $[SS]_c + 2[SI]_c + [II]_c = nN$ azonossággal, a következőt kapjuk

$$\gamma nNU = \gamma Z^2 + ZU(\tau + 2\gamma) + \gamma U^2, \quad (5.20)$$

ami egy másodfokú egyenlet Z -re az U függvényében, és könnyen látható, hogy minden $U \in [0, nN]$ esetén egyetlen nemnegatív Z megoldása van, melyet $Z = g(U)$ -val jelölünk a továbbiakban. Kiszámítható továbbá, hogy $g(0) = 0$ és $g(nN) = 0$.

Definiáljuk az f függvényt a következő formulával:

$$f(U) := \frac{\tau U}{U + g(U)} \sum_{l=1}^L \frac{n_l(n_l - 1)N_l}{\gamma(U + g(U)) + \tau n_l g(U)} = \frac{\tau U}{(U + g(U))^2} \sum_{l=1}^L \frac{n_l(n_l - 1)N_l}{\gamma + \tau n_l \frac{g(U)}{U + g(U)}}.$$

Az endemikus egyensúlyi állapot egyértelmű létezése ekvivalens azzal, hogy egyetlen olyan $U \in (0, nN)$ létezik, melyre $f(U) = 1$. Ezt három lépésben fogjuk belátni: egyrészt megnézzük hogy az f függvény határértéke 0-ban kisebb, mint 1, másrészt a függvényérték nN -ben nagyobb, mint 1, harmadrészt f a $(0, nN)$ intervallumon szigorúan monoton növekvő, folytonos függvény.

Mivel f a 0-ban nem definiált, ezért a határértékét tekintjük 0-ban. Az (5.20) egyenletet átalakíthatjuk:

$$\gamma nNU = \gamma(Z + U)^2 + \tau ZU,$$

amiből

$$\frac{U}{(U + Z)^2} = \frac{\gamma}{\gamma nN - \tau Z},$$

így

$$\lim_{U \rightarrow 0} \frac{U}{(U + g(U))^2} = \frac{1}{nN}.$$

Elosztva Z -vel az (5.20) egyenletet látható, hogy

$$\lim_{U \rightarrow 0} \frac{U}{g(U)} = 0, \text{ így } \lim_{U \rightarrow 0} \frac{g(U)}{U + g(U)} = 1,$$

majd felhasználva a kapott határértékeket

$$\lim_{U \rightarrow 0} f(U) = \frac{\tau}{nN} \sum_{l=1}^L \frac{n_l(n_l - 1)N_l}{\gamma + \tau n_l} < \frac{\tau}{nN} \sum_{l=1}^L \frac{n_l(n_l - 1)N_l}{\tau n_l} = \frac{N(n-1)}{Nn} < 1.$$

Az f függvényértékét az nN helyen könnyedén kiszámolhatjuk:

$$f(nN) = \frac{\tau}{nN} \sum_{l=1}^L \frac{n_l(n_l - 1)N_l}{\gamma} = \frac{\tau}{\gamma} \frac{\langle n^2 \rangle - \langle n \rangle}{\langle n \rangle},$$

ebből látszik, hogy ha $\tau_c^{CPW} < \tau$, akkor $f(nN) > 1$.

Az f függvény folytonossága a $(0, nN)$ intervallumon látszik, hiszen ott g folytonos, így már csak az f függvény szigorú monoton növekedését kell igazolni ezen az intervallumon, amihez elegendő lenne az összeg tagjairól külön-külön belátni, hogy szigorúan monoton növekszenek, azaz az

$$U \mapsto \frac{\tau U}{U + g(U)} \frac{n_l(n_l - 1)N_l}{\gamma(U + g(U)) + \tau n_l g(U)}$$

függvényekről, minden $l = 1, \dots, L$ esetén. Ezek a függvények pontosan akkor monotonok, ha az

$$U \mapsto \frac{U}{U + g(U)} \frac{1}{\gamma(U + g(U)) + \tau n_l g(U)}$$

függvények is. Apróbb átalakítások után és felhasználva a fentebb említett $\gamma(U + g(U))^2 = \gamma nNU - \tau U g(U)$ összefüggést

$$\begin{aligned} \frac{U}{U + g(U)} \frac{1}{\gamma(U + g(U)) + \tau n_l g(U)} &= \frac{U}{(\gamma + \tau n_l)(U + g(U))^2 - \tau n_l U(U + g(U))} = \\ &= \frac{U}{\frac{\gamma + \tau n_l}{\gamma}(\gamma nNU - \tau U g(U)) - \tau n_l U(U + g(U))} = \\ &= \frac{U}{U((\gamma + \tau n_l)nN - \tau n_l U) - U g(U)(\tau(1 + \frac{\tau}{\gamma} n_l + n_l))}. \end{aligned}$$

Tehát elegendő lenne az $U \mapsto \frac{U}{U((\gamma + \tau n_l)nN - \tau n_l U) - U g(U)(\tau(1 + \frac{\tau}{\gamma} n_l + n_l))}$ függvény szigorú monoton növekedése helyett belátni, hogy annak reciproka, a

$$U \mapsto (\gamma + \tau n_l)nN - \tau n_l U - g(U)(\tau(1 + \frac{\tau}{\gamma} n_l + n_l))$$

függvény szigorúan monoton csökken az $U \in (0, nN)$ intervallumon, vagyis elhagyva illetve leosztva a felesleges pozitív konstansokkal, vizsgáljuk meg, hogy a

$$h_l(U) := U + g(U)\left(\frac{1}{n_l} + \frac{\tau}{\gamma} + 1\right)$$

függvény szigorúan növekedő-e. Ez következne abból a tényből, hogy a h derivált függvénye szigorúan pozitív a kérdéses intervallumon, azaz $h'_l(U) = 1 + g'(U)\left(\frac{1}{n_l} + \frac{\tau}{\gamma} + 1\right) > 0$, ha $U \in (0, nN)$. Ennek igazolásához meg kell vizsgálni a g függvényt, ami az (5.20) másodfokú egyenlet egyértelmű nemnegatív megoldása az U változó mellett, tehát képlettel is megadható:

$$g(U) = \frac{1}{2\gamma} \left(-U(\tau + 2\gamma) + \sqrt{U^2(\tau^2 + 4\gamma\tau) + 4UnN\gamma^2} \right).$$

Illetve a g függvény első és második deriváltja is kiszámítható:

$$g'(U) = \frac{1}{2\gamma} \left(-(\tau + 2\gamma) + \frac{2U(\tau^2 + 4\gamma\tau) + 4nN\gamma^2}{\sqrt{U^2(\tau^2 + 4\gamma\tau) + 4UnN\gamma^2}} \right),$$

$$g''(U) = \frac{-2n^2N^2\gamma^3}{(U^2(\tau^2 + 4\gamma\tau) + 4UnN\gamma^2)^{\frac{3}{2}}} < 0.$$

Azaz g szigorúan konkáv függvény a $(0, nN)$ intervallumon, vagyis g' derivált függvény szigorúan monoton csökken és könnyen kiszámítható, hogy $g'(nN) = -\frac{\gamma}{\tau + 2\gamma}$. Így minden $U \in (0, nN)$ mellett $g'(U) > -\frac{\gamma}{\tau + 2\gamma}$. Ezen kívül látható, hogy $\lim_{U \rightarrow 0} g'(U) = +\infty$, így a g' függvénynek pontosan egy zérushelye van a $(0, nN)$ intervallumon. Azon az intervallumon, ahol $g' \geq 0$, ott természetesen $h'_l > 0$, tehát csak azt az intervallumot kell megnézni, ahol $g' < 0$.

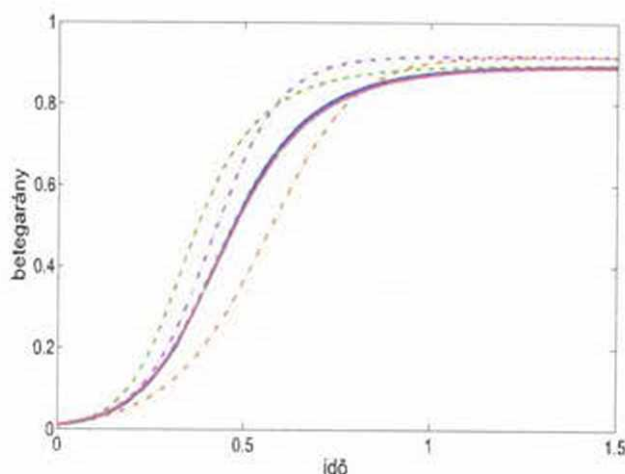
Vizsgáljuk meg az $1 + g'(U)\left(1 + \frac{\tau}{\gamma} + 1\right)$ kifejezést tetszőleges $U \in (0, nN)$ mellett, kihasználva a g' monotonitását:

$$1 + g'(U)\left(1 + \frac{\tau}{\gamma} + 1\right) > 1 - \frac{\gamma}{\tau + 2\gamma}\left(2 + \frac{\tau}{\gamma}\right) = 0.$$

Azon az intervallumon, ahol $g' < 0$, fennáll:

$$h'_l(U) = 1 + g'(U)\left(\frac{1}{n_l} + \frac{\tau}{\gamma} + 1\right) > 1 + g'(U)\left(1 + \frac{\tau}{\gamma} + 1\right) > 0,$$

vagyis h'_l pozitív azon az intervallumon is, ahol $g' < 0$. Tehát $h'_l(U) > 0$, minden $U \in (0, nN)$ mellett, így ezzel igazoltuk az f függvény szigorú monotonitását a $(0, nN)$ intervallumon. \square



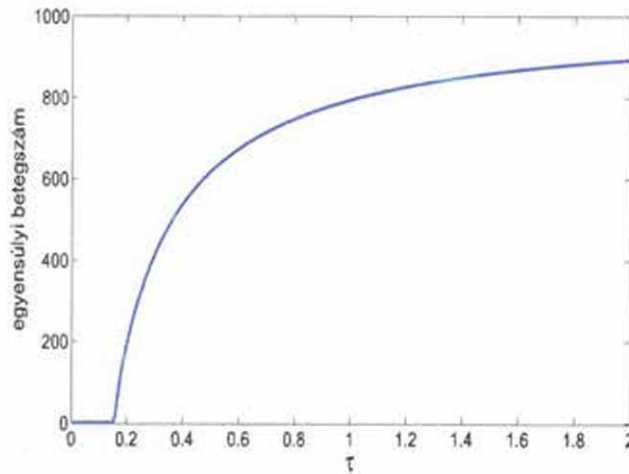
5.1. ábra. Az átlagos betegszám arány az idő függvényében bimodális gráfon az (5.4) (lila szaggatott görbe), az (5.5) (barna szaggatott görbe), az (5.8) (zöld szaggatott görbe), az (5.9) (cián szaggatott görbe) és az (5.16)-(5.19) egyenletek (kék görbe) esetén illetve a szimulációból (piros görbe). $N = 1000$, $N_3 = 500$, $N_9 = 500$, $\tau = 2$, $\gamma = 1$. 250 szimuláció átlagából, 10 kezdeti beteggel.

5.4. Összefoglalás

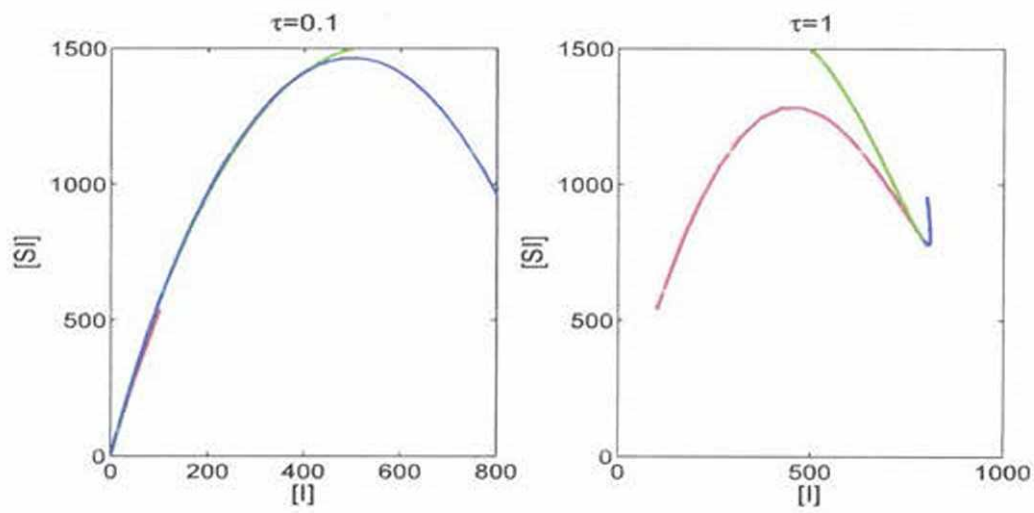
A fejezetben ismertettünk öt alacsony dimenziós közelítő differenciálegyenlet-rendszert *SIS* típusú járványterjedés modellezésére. Először a különböző módszereket hasonlítottuk össze egy könnyen kezelhető heterogén fokszámeloszlású gráfon. A szimulációból kimért betegszám arányt vetettük össze a különböző eljárások eredményeivel az 5.1. ábrán bimodális gráf esetén. Látható, hogy a legjobb közelítést az (5.9) és a (5.16)-(5.19) modellek adják, ami összhangban áll a fenti gondolatokkal, miszerint heterogén gráfon ezek a legjobb alacsony dimenziós módszerek az öt eljárás közül. Fontos megjegyezni, hogy a csökkentett méretű (5.16)-(5.19) egyenletrendszer még jobb közelítést is ad, mint a nagy dimenziós (5.9) rendszer.

Az (5.4), az (5.5), az (5.8) és az (5.9) modellek bemutatása után, a (5.16)-(5.19) egyenletek vizsgálata következett, amely rendszernek ismertetése megtalálható az irodalomban [16], azonban analitikus vizsgálata még nem teljes, az endemikus egyensúlyi pont egyértelműségének igazolása a fejezet új eredménye. Az elméleti állítást a numerikus eredmények is jól mutatják. Az 5.2. ábrán a betegek átlagos számának egyensúlyi értékét - azaz az $[I] = \sum_{l=1}^L [I_l]$ értéket az endemikus egyensúlyi állapotban - követhetjük nyomon a τ paraméter függvényében, rögzített

$\gamma = 1$ mellett, ekkor $\tau_c^{CPW} = 0.1538$. Az (5.16)-(5.19) rendszert itt is egy 1000 csúcsú bimodális gráf esetén oldottuk meg, illetve az 5.3. ábrán is ilyen gráftípust használtunk. Az 5.3. ábra baloldalán $\gamma = 1$ és $\tau = 0.1$ esetén ábrázoltuk a fázisképet az $[I] - [SI]$ síkon, ilyenkor a triviális egyensúlyi pont stabil, így minden megoldás a $(0, 0)$ pontba tart. Az 5.3. ábra jobboldalán $\gamma = 1$ és $\tau = 1$ esetén láthatjuk a fázisképet az $[I] - [SI]$ síkon. Ezen értékek mellett $\tau_c^{CPW} < \tau$, tehát a triviális egyensúlyi pont instabil és a numerikus eredményekből azt sejtethetjük, hogy az endemikus egyensúlyi helyzet stabil lesz. Ennek igazolása képezi további kutatásunk tárgyát ebben a témában. Emellett az egyensúlyi pontok globális stabilitásának kérdésével is érdemes a jövőben foglalkozni. Azaz szeretnénk megvizsgálni, hogy $\tau < \tau_c^{CPW}$ esetén minden megoldás a triviális egyensúlyi helyzetbe tart-e, illetve $\tau_c^{CPW} < \tau$ mellett pedig az endemikus egyensúlyi pontba, ahogy azt az ábrák sugallják.



5.2. ábra. A betegek átlagos számának egyensúly értéke τ függvényében bimodális gráfon az (5.16)-(5.19) egyenletek megoldása alapján. $N = 1000$, $N_3 = 500$, $N_9 = 500$, $\gamma = 1$, $\tau_c^{CPW} = 0.1538$.



5.3. ábra. Az (5.16)-(5.19) egyenlet egy fázisképe: $[SI]$ értékei az $[I]$ átlagos betegszám függvényében bimodális gráfon $\tau = 0.1$ és $\tau = 1$ esetén. 100 (piros), 500 (kék) és 800 (zöld) kezdeti betegből indítva. $N = 1000$, $N_3 = 500$, $N_9 = 500$, $\gamma = 1$, $\tau_c^{CPW} = 0.1538$.

6. fejezet

Hálózati modell együttes csődvalószínűségek meghatározására

Ebben a fejezetben a hálózati modellek valós életben történő alkalmazhatóságára adunk példát egy, a pénzügyi életben megjelenő probléma új típusú megközelítésével. A problémával Bihary Zsolt (Morgan Stanley) felvetésére kezdtünk el foglalkozni és vele együttműködve végeztük a vizsgálatokat. Ennek a résznek az is az újdonsága az eddigi munkákhoz képest, hogy egy súlyozott gráfon SI dinamikát fogunk vizsgálni, azaz a folyamat egyirányú, illetve - az előbbi fejezetekhez képest - másfajta átmeneti ráták használatára lesz szükség, ami érzékelteti a hálózatokon zajló folyamatok sokszínűségét.

6.1. Motiváció

A pénzügyi piacokon a befektetők számára lehetőség van arra, hogy például kötvénybefektetések esetén bizonyos hitelderivatívák segítségével a kötvény kibocsátójának nemfizetési kockázata ellen védjék magukat. Tehát ezekre a hitelbiztosításokra abban az esetben van szükség, ha a kötvénykibocsátó cég csődbe menne a futamidő lejáratá előtt. Az egyik legegyszerűbb a mulasztási csereügylet, azaz CDS (credit default swap), ami védelmet nyújt egy kockázatos kötvény vétele esetén. Ennél bonyolultabb biztosítás a basket CDS, azaz BDS (basket default swap), ami már több kötvény esetén nyújt fedezetet. A fejezetben feltesszük, hogy különböző cégek bocsátják ki a különböző kötvényeket. Aszerint rangsoroljuk a BDS-eket, hogy hány kötvénykibocsátó cég csődbe menetele után fizet kártérítést: first-to-default (FTD) esetén az

első csődbe kerülést állapítják meg kifizetési időpontnak, míg például az n -th-to-default (NTD) azt jelenti, hogy a hitelbiztosítás kiírója az n -edik cégbedőlés után fizet csak. A BDS biztosításokat kibocsátó intézményeknél fontos kérdés, hogy ezeket a hitelbiztosításokat hogyan lehet arbitrázs-mentesen beárazni, azaz félreárazás nélkül. A BDS-ek értékét az egyes cégek együttes csődvalószínűségei alapján határozzák meg. Erre többek között a Gauss kopulás módszert használják [19], ami egy heurisztikus, valószínűségszámításon alapuló eljárás. Az értekezésben alternatív módszereket mutatunk be, melyek a cégek együttes csődbe jutási valószínűségeit adják meg a piac jelenlegi helyzete alapján, illetve a hálózat felépítésének függvényében. Ezek az új eljárások hálózati modelleken alapulnak és versenyképes alternatívát jelenthetnek a Gauss kopulás módszerrel szemben, illetve előnyösebbek is bizonyos értelemben, ugyanis betekintést engednek az egyes cégek állapotaiba bármely időpillanatban, emellett összetettebb pénzügyi rendszerek modellezésére is alkalmasak. Az új modellekkel szerzett többlet-információ lehetőséget teremt arra, hogy egyrészt vizsgáljuk a kötvények együttes bedőlési valószínűségeit, másrészt a pénzügyi élet egy másik fontos kérdésére is megpróbáljunk választ találni: hogyan állítsuk be a cégek egymás közötti erőviszonyát és hogyan építsük fel a hálózatot úgy, hogy a rendszerkockázat, vagyis a csődben lévő cégek számának várható értéke minimális legyen.

6.2. A probléma felvázolása és a Gauss kopulás módszer bemutatása

Legyen adva N darab kötvénykibocsátó cég és tegyük fel, hogy ismerjük valamely T időpontban az egyes cégek csődbe jutási valószínűségeit, jelöljük ezeket \bar{p}_q -val, $q = 1, \dots, N$, illetve adott még a cégek közötti kapcsolat erőssége is. A kérdés, hogy hogyan lehet ez alapján kiszámítani az együttes csődesélyeket, azaz a $D(i)$ értékeket, amik azt adják meg, hogy mi a valószínűsége, hogy pontosan i darab cég van csődben a T időpontban, $i = 0, \dots, N$. Ezen $D(i)$ értékek alapján történik a BDS-ek beárazása, de ennek az eljárásnak az ismertetése nem képezi az értekezés tárgyát, számunkra csak a $D(i)$ értékek meghatározása a cél. Először a pénzügyi életben használt Gauss kopulás módszert fogjuk röviden bemutatni, majd összevetjük az eredményeit az általunk kínált modellel és megvizsgáljuk, hogy a két eljárást egymáshoz lehet-e kalibrálni a különböző bemeneti paraméterek megfelelő beállításával.

Jelölje $V_q(t)$ standard normális eloszlású valószínűségi változó a q -adik cég kötvényének

árfolyamát a t időpontban és legyen $B_q(t)$ egy determinisztikus korlát, $q = 1, \dots, N$, $t \in \mathbb{R}_0^+$. Ha az árfolyam a B_q korlát alá esik, akkor a cég csődbe megy végleg, azaz $p_q(t)$, a q -adik cég csődjének a valószínűsége t -ben:

$$p_q(t) = P(V_q(t) < B_q(t)) = \phi(B_q(t)),$$

ahol ϕ a standard normális eloszlás eloszlásfüggvénye.

A kötvények árfolyamai közötti kapcsolat a következőképpen van megadva, rögzített $\rho \in [0, 1]$ paraméter mellett:

$$V_q = \sqrt{\rho}Z + \sqrt{1-\rho}Y_q, \quad q = 1, \dots, N$$

ahol Z és Y_q , $q = 1, \dots, N$ független standard normális eloszlású valószínűségi változók. Ezek alapján $V_q \sim N(0, 1)$ és könnyen látható, hogy bármely két cég kötvényének az árfolyama közötti korreláció ρ , tehát a ρ értéke adja meg a cégek egymásra hatásának erősségét, ami ebben a modellben minden cég között ugyanakkora, tehát a kapcsolati háló homogén. Emellett látható, hogy a cégek kötvényeinek árfolyamát nem csak a többi céggel való kapcsolat befolyásolja, hanem a Z valószínűségi változó is, ami így meghatározza a cégek spontán csődbe jutásának esélyét.

Tehát adottak valamely T időpontban az egyes cégek csődjének a valószínűségei, azaz a $\tilde{p}_q = p_q(T)$ értékek, $q = 1, \dots, N$, illetve $\rho \in [0, 1]$. A $B_q(T)$ korlátok a T időpontban könnyen visszaszámolhatók a bemeneti adatok alapján: $B_q(T) := \phi^{-1}(\tilde{p}_q)$. Mivel a V_q valószínűségi változók nem függetlenek, így a $D(i)$ értékek nehezen számolhatóak ki. Innen jön az ötlet, hogy rögzítsünk le egy konkrét $z \in \mathbb{R}$ értéket, és vizsgáljuk $Z = z$ esetén a különböző feltételes valószínűségeket, ugyanis ebben az esetben a V_q változók függetlenek. Ekkor $V_q = \sqrt{\rho}z + \sqrt{1-\rho}Y_q$, vagyis $V_q \sim N(\sqrt{\rho}z, \sqrt{1-\rho})$ és

$$p_q(T|Z = z) = P(V_q(T) < B_q(T)|Z = z) = \phi\left(\frac{B_q(T) - \sqrt{\rho}z}{\sqrt{1-\rho}}\right),$$

ahol $p_q(T|Z = z)$ jelöli a q -adik cég csődjének a valószínűségét t -ben, feltéve, hogy $Z = z$. Ezen értékek függvényében kiszámíthatók a $D(i|Z = z)$ mennyiségek, azaz annak a valószínűsége, hogy pontosan i darab cég van csődben T -ben, feltéve, hogy a Z valószínűségi változó értéke a rögzített z szám. Itt fogjuk kihasználni, hogy mivel a V_q változók függetlenek, így a különböző cégek csődben levésének eseményei függetlenek, tehát szorzatként felírhatóak. Például annak az esélye a $Z = z$ feltétel mellett, hogy 0 darab cég van csődben:

$D(0|Z = z) = \prod_{q=1}^N (1 - p_q(T|Z = z))$, illetve a többi $D(i|Z = z)$ feltételes valószínűség is hasonlóan számolható.

Ezek után használjuk a teljes valószínűség tételét folytonos esetben, így megkapjuk a keresett $D(i)$ értékeket, azaz hogy mi a valószínűsége annak, hogy pontosan i cég van csődbe a T időpontban:

$$D(i) = \int_{-\infty}^{+\infty} D(i|Z = z)\varphi(z)dz, \quad i = 0, \dots, N,$$

ahol φ a standard normális eloszlás sűrűségfüggvénye. Érdekes észrevétel, hogy habár a kimeneti adatokat is abban az időpontban kapjuk meg, mint amiben a bemeneti adatokat adtuk meg, a módszer során nem kellett ismerni a T időpontot. Ez a Gauss kopulás módszer rövid összefoglalása, amit majd összehasonlítunk az következő fejezetben ismertetésre kerülő hálózatos modellel.

6.3. A hálózati megközelítésen alapuló modell

Most nézzük meg, hogyan lehet ezt a problémát elhelyezni a hálózati modellek világában. Egy természetesen adódó hálózati modellt ajánlunk a cégek együttes csődvalószínűségeinek leírására. Ez a modell alkalmas a Gauss kopulás módszerrel való összevetésre, illetve a rendszerkockázat vizsgálatára, azonban csak kis N -re alkalmazható a nagy számítási igény miatt. A 6.4. fejezetben egy másik módszert kínálunk, ami nagy N -re is alkalmazható.

6.3.1. A hálózati modell leírása

Egy súlyozott gráffal fogjuk szemléltetni a cégek közötti kapcsolati hálót. A gráf csúcsait tekintjük a cégeknek, míg az egyes csúcsok közötti éleken lévő súly jelzi az adott cégek közötti kapcsolat erősségét. Legyen N a csúcsok száma és bármely csúcs két lehetséges állapotban lehet: S -működő, I -csődbe ment, azaz járványterjedési folyamatként tekintünk a cégek csődbe jutására. Így egy 2^N állapotterű $\mathbf{X}(t)$, $t \in \mathbb{R}_0^+$ Markov-lánchoz jutunk, ahol $\mathbf{X}(t)$ egy olyan valószínűségi változó, mely minden rögzített $t \in \mathbb{R}_0^+$ időpontban megadja, hogy a folyamat melyik állapotban van. Az állapotteret a szokásos \mathcal{S} halmaz adja, és vezessük be az \mathcal{S}^k jelölést a k darab csődből tartalmazó állapotok részhalmazára, így \mathcal{S}^k elemei: $\{\mathcal{S}_1^k, \mathcal{S}_2^k, \dots, \mathcal{S}_{d_k}^k\}$, ahol $d_k = \binom{N}{k}$, $k = 0, \dots, N$. Az \mathcal{S}^k halmaz i -edik eleme esetén az \mathcal{S}_i^k állapotban az l -edik csúcs állapotát $\mathcal{S}_i^k(l)$ jelöli.

Feltesszük, hogy kezdetben minden cég működik, és a csődbe menés végleges állapotváltozás, azaz bármely csúcs esetén csak $S \rightarrow I$ átmenet történhet, tehát SI típusú dinamikával dolgozunk. Bármely cég csődbe jutását két tényező befolyásolja: egyrészt minden céghez tartozik egy $\lambda_q \in \mathbb{R}_0^+$, $q = 1, \dots, N$ spontán csődbe menési ráta, ami a cégek sérülékenységet mér, másrészt a már csődbe jutott cégek hatnak a még működő szomszédokra a gráf $G = ((\omega_{p,q})) \in \mathbb{R}^{N \times N}$ adjacencia mátrixában tárolt nemnegatív rátákkal, tehát $\omega_{p,q}$ jelentse a p -edik cég hatáserősségét a q -adik cégre, ezt a hatáserősséget élsúlynak is nevezzük a továbbiakban. Az $S \rightarrow I$ átmenetet egy exponenciális eloszlású valószínűségi változó írja le, azaz csak az eloszlás paraméterét, az f_{SI} rátát kell megadni. Legyen

$$f_{SI}(l, G, S_i^k) := \lambda_l \left(1 + \sum_{\{p \in \{1, \dots, N\} | S_i^k(p) = I\}} \omega_{p,l} \right), \quad (6.1)$$

feltéve, hogy $S_i^k(l) = S$, azaz az l -edik cég még nincs csődbe. Ezt az átmeneti rátát Bihary Zsolt ajánlására választottuk és a későbbiekben látni fogjuk, hogy hatékonyan írja le a folyamat működését.

A hálózat és a dinamika leírása után felírjuk a folyamathoz kapcsolódó 2^N dimenziós (2.2) alapegyenlet-rendszert. Ehhez vezessük be minden $S_i^k \in \mathcal{S}^k$, $i = 1, \dots, d_k$, $k = 0, \dots, N$ esetén a $H_{S_i^k}^-$, az " S_i^k múltja" és a $H_{S_i^k}^+$, az " S_i^k jövője" halmazokat:

$$H_{S_i^k}^- := \left\{ h \in \{1, \dots, d_{k-1}\} \mid \right.$$

$$\left. \exists! l_h \in \{1, \dots, N\} : S_i^k(l_h) = I, S_h^{k-1}(l_h) = S, S_i^k(m) = S_h^{k-1}(m), \forall m \neq l_h, m = 1, \dots, N \right\},$$

$$H_{S_i^k}^+ := \left\{ j \in \{1, \dots, d_{k+1}\} \mid \right.$$

$$\left. \exists! l_j \in \{1, \dots, N\} : S_i^k(l_j) = S, S_j^{k+1}(l_j) = I, S_i^k(m) = S_j^{k+1}(m), \forall m \neq l_j, m = 1, \dots, N \right\}.$$

Tehát az S_i^k múltja halmaz tartalmazza az összes olyan \mathcal{S}^{k-1} halmazbeli állapot indexét, melyek pontosan egy csúcs állapotában térnek el az S_i^k állapottól, hiszen csak ilyen állapotok között következhet be közvetlen állapotváltozás. Az S_i^k jövője halmaz jelentése is értelemszerűen adódik az előbbi alapján.

Jelölje $x_{S_i^k}(t)$ annak a valószínűségét, hogy a t időpontban a Markov-folyamat az S_i^k állapotban van. Ekkor a (2.2) alapegyenlet-rendszer alakja:

$$\dot{x}_{S_i^k}(t) = \sum_{h \in H_{S_i^k}^-} \lambda_{l_h} \left(1 + \sum_{\{p \in \{1, \dots, N\} | S_h^{k-1}(p) = I\}} \omega_{p,l_h} \right) x_{S_h^{k-1}}(t) - \quad (6.2)$$

$$\left(\sum_{j \in H_{S_i^k}^+} \lambda_{l_j} (1 + \sum_{\{p \in \{1, \dots, N\} | S_i^k(p) = I\}} \omega_{p, l_j}) \right) x_{S_i^k}(t), \quad i = 1, \dots, d_k, \quad k = 0, \dots, N.$$

Az egyenletrendszerre a kompaktabb $\dot{x}(t) = Px(t)$ alakban is hivatkozhatunk.

Az alapegyenletek felírását egy 3 csúcsú gráfon szemléltetjük.

13. Példa. Legyen $N = 3$. $A = ((\omega_{p,q})) \in \mathbb{R}^{3 \times 3}$ a gráf adjacencia mátrixa és a λ_q , $q = 1, 2, 3$ értékek a spontán csődbe kerülés rátái. Ekkor az alapegyenletek:

$$\begin{aligned} \dot{x}_{SSS}(t) &= -(\lambda_1 + \lambda_2 + \lambda_3)x_{SSS}(t), \\ \dot{x}_{ISS}(t) &= \lambda_1 x_{SSS}(t) - (\lambda_2(1 + \omega_{1,2}) + \lambda_3(1 + \omega_{1,3}))x_{ISS}(t), \\ \dot{x}_{SIS}(t) &= \lambda_2 x_{SSS}(t) - (\lambda_1(1 + \omega_{2,1}) + \lambda_3(1 + \omega_{2,3}))x_{SIS}(t), \\ \dot{x}_{SSI}(t) &= \lambda_3 x_{SSS}(t) - (\lambda_1(1 + \omega_{3,1}) + \lambda_2(1 + \omega_{3,2}))x_{SSI}(t), \\ \dot{x}_{IIS}(t) &= \lambda_1(1 + \omega_{2,1})x_{SIS}(t) + \lambda_2(1 + \omega_{1,2})x_{ISS}(t) - \lambda_3(1 + \omega_{1,3} + \omega_{2,3})x_{IIS}(t), \\ \dot{x}_{ISI}(t) &= \lambda_1(1 + \omega_{3,1})x_{SSI}(t) + \lambda_3(1 + \omega_{1,3})x_{ISS}(t) - \lambda_2(1 + \omega_{1,2} + \omega_{3,2})x_{ISI}(t), \\ \dot{x}_{SII}(t) &= \lambda_2(1 + \omega_{3,2})x_{SSI}(t) + \lambda_3(1 + \omega_{2,3})x_{SIS}(t) - \lambda_1(1 + \omega_{2,1} + \omega_{3,1})x_{SII}(t), \\ \dot{x}_{III}(t) &= \lambda_3(1 + \omega_{1,3} + \omega_{2,3})x_{IIS}(t) + \lambda_2(1 + \omega_{1,2} + \omega_{3,2})x_{ISI}(t) + \lambda_1(1 + \omega_{2,1} + \omega_{3,1})x_{SII}(t). \end{aligned} \quad (6.3)$$

A rendszer második egyenlete például azt rögzíti, hogy az (ISS) állapotba az (SSS) állapotból juthatunk el spontán csődbemenéssel λ_1 rátával, emellett az (ISS) állapotból kimehetünk $\lambda_2(1 + \omega_{1,2})$ rátával az (IIS) állapotba, úgy hogy az első cég csődbe viszi a másodikat, illetve az (ISS) állapotból átmehetünk az (ISI) állapotba, úgy hogy az első cég csődbe juttatja a harmadikat $\lambda_3(1 + \omega_{1,3})$ rátával.

6.3.2. Eljárás a csődvalószínűségek meghatározására a hálózati modell segítségével

Most megmutatjuk, hogy ez a hálózati modell hogyan használható a fent ismertetett árazási probléma kezelésére, azaz hogyan lehet eljutni az egyes cégek csődbe menési valószínűségeinek ismeretéből az együttes bedőlési valószínűségekhez. A Gauss kopulás módszerben a cégek közötti viszonyt a ρ paraméter adja meg, míg a hálózati modellnél a kapcsolat erősségét a G szomszédsági mátrix írja le, jelen esetben $\omega_{q,p} := \omega$, valamely rögzített ω esetén, $p, q = 1, \dots, N$, így megtartva a homogén kapcsolati hálót. Tehát adott a cégek közötti kapcsolat erőssége ω

paraméterrel, illetve adottak a T időpontban az egyes cégek csődbe jutási valószínűségei, azaz a \tilde{p}_q értékek, $q = 1, \dots, N$. Újdonság, hogy ennél a módszernél a T időpontot is pontosan meg kell adni, ellentétben a Gauss kopulás eljárással.

Tekintsünk a (6.2) alapegyenlet-rendszert. A $p_q(t)$ valószínűségek és az egyenletrendszer változói között felírható egy egyszerű összefüggés minden $t \in \mathbb{R}_0^+$ esetén:

$$p_q(t) = \sum_{k=1}^N \sum_{\{i \in \{1, \dots, d_k\} : S_i^k(q) = I\}} x_{S_i^k}(t), \quad q = 1, \dots, N. \quad (6.4)$$

Ezek a képletek azt fejezik ki, hogy a q -adik cég pontosan akkor van csődbe, ha a folyamat egy olyan állapotban van, ahol a q -adik cég csődbe van, így ezek diszjunkt uniója kiadja a vizsgált eseményt, azaz a valószínűségek összege megadja a p_q értéket.

Fontos látni, hogy a hálózati modellhez szükséges λ_q spontán bedőlési ráták nincsenek megadva, így először szeretnénk megkapni őket az eddigi információk alapján. Ehhez bevezetjük a következő \mathfrak{F} függvényt:

$$\begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_N \end{pmatrix} \rightarrow \begin{pmatrix} p_1(T) - \tilde{p}_1 \\ \vdots \\ p_N(T) - \tilde{p}_N \end{pmatrix},$$

ahol $p_q(T)$ jelöli a (6.4) összefüggéssel megadott függvényeket a T időpontban. Először megoldjuk numerikusan a (6.3) egyenletrendszert tetszőleges $(\lambda_1, \dots, \lambda_N)$ vektor és a kezdetben megadott ω bemeneti paraméter mellett, az $x(0) = (1, 0, \dots, 0)$ kezdeti feltétellel, azaz a 0 időpontban minden cég működik még. Ezek után kiszámítjuk az egyenletrendszer megoldásainak felhasználásával a (6.4)-ben megadott p_q valószínűségeket a T időpontban. Azt keressük, hogy mely λ_q értékek mellett vesz fel az \mathfrak{F} függvény nullát. A numerikus tapasztalatok azt mutatják, hogy ezen λ_q értékek egyértelműen meghatározhatóak.

Tehát visszakerestük azokat a λ_q , $q = 1, \dots, N$ értékeket, melyek esetén a T időpontban az egyes cégek bedőlési valószínűségei megegyeznek az előre elvárt \tilde{p}_q értékekkel, $q = 1, \dots, N$. Miután megvannak a megfelelő spontán bedőlési ráták értékei, numerikusan megoldható a (6.2) alapegyenlet-rendszer és kiszámíthatóak a megoldásfüggvények segítségével az együttes bedőlések valószínűségei:

$$D_i(t) = \sum_{j=1}^{d_j} x_{S_j^i}(t), \quad (6.5)$$

ahol $D_i(t)$ jelöli annak a valószínűségét, hogy i darab cég van csődbe a t időpontban.

Az eljárás bemutatása után, szeretnénk a Gauss kopulás és a hálózatos módszerek eredményeit összevetni azonos bemeneti paraméterek esetén, azonban ez nehézségbe ütközik, mert a cégek közötti kapcsolatokért felelős mennyiségek, a ρ és az ω viszonya nem ismert. Azonban ezen bemeneti paraméterek megfelelő választásával a két eljárás egymáshoz kalibrálható, azaz a hálózatos modellel kapott $D_i(T)$ értékek jól illeszkednek a Gauss kopulás módszer $D(i)$ értékeire, amit a 6.1. ábrán szemléltetünk. Itt az látható, hogy előre megadott \tilde{p}_q valószínűségek esetén a ρ paraméterhez megválasztható egy olyan ω érték, mely esetén a $D(i)$ és a $D_i(T)$ valószínűségek megfelelően közel vannak egymáshoz. Ez azt mutatja, hogy ez az új módszer is jól használható a feladat kezelésére. A mi módszerünk előnye, hogy megadja a folyamat összes állapotának a valószínűségét minden időpillanatban, szemben a Gauss kopulás módszerrel, ami csak az együttes csődvalószínűségeket határozza meg és csak a vizsgált T időpontban. Ez az új eljárás így lehetőséget teremt a folyamat mélyrehatóbb vizsgálatára, illetve heterogén kapcsolati hálóval ellátott hálózat elemzését is lehetővé teszi.

A fejezet további részében egy új problémával foglalkozunk, kihasználva a hálózati modell nyújtotta előnyöket. A modell eredményeként kapott $D_i(t)$ függvények segítségével a pénzügyi élet egy fontos kérdésére is választ találhatunk: hogyan változik a rendszerkockázat az idő függvényében különböző kapcsolati hálók esetén. A rendszerkockázatot a

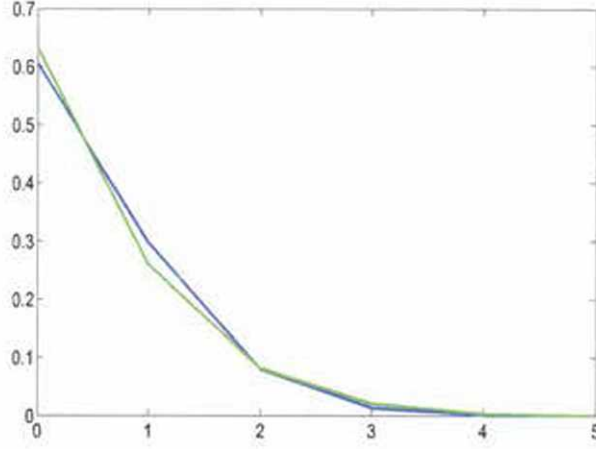
$$K(t) = \sum_{i=1}^N i D_i(t),$$

formulával definiálva, ami a csődök számának várható értékét adja meg t időpontban, már vizsgálhatóvá válik ez a kérdés is.

Sajnos, az ajánlott új eljárás - és így a rendszerkockázat meghatározása - csak kis N esetén működik, ugyanis beleütközünk a szokásos gondba, hogy a 2^N méretű alapegyenlet-rendszer túl nagy számítási igényű. A következő szakaszban egy másik eljárást kínálunk a rendszerkockázat mérésére, mely nagy N -re is alkalmazható.

6.4. A csúcsok szintjén felírt egyenletek

A (6.2) alapegyenlet-rendszer egyenletei alapján szeretnénk felírni egy kisebb dimenziós közelítő differenciál-egyenletrendszert, ami már nagy N -re is numerikusan megoldható. A redukált rendszer segítségével kiszámoljuk a rendszerkockázatot, illetve azt tanulmányozzuk, hogy



6.1. ábra. A két módszerrel kapott $D(i)$ és $D_i(T)$ értékek, $i = 0, \dots, 5$ összehasonlítása (zöld görbe: Gauss Kopula, kék görbe: exact modell). Bemeneti adatok: $(\tilde{p}_1, \tilde{p}_2, \tilde{p}_3, \tilde{p}_4, \tilde{p}_5) = 0.1014 \cdot (1, 1, 1, 1, 1)$, $\omega = 0.35$, $\rho = 0.2$ $T = 1$.

hogyan befolyásolja a gráf struktúrája a rendszerkockázat alakulását különböző, nagy méretű hálózatok esetén.

6.4.1. A modell formalizálása

Vezessünk be néhány új jelölést. Legyen $\langle I_q \rangle(t)$ annak a valószínűsége, hogy a q -adik cég csődben van a t időpontban. Tehát

$$\langle I_q \rangle(t) := \sum_{k=1}^N \sum_{\{i \in \{1, \dots, d_k\} | S_i^k(q) = I\}} x_{S_i^k}(t), \quad q = 1, \dots, N.$$

Emellett használni fogjuk a következő jelöléseket is:

$$\langle S_q \rangle(t) := \sum_{k=0}^{N-1} \sum_{\{i \in \{1, \dots, d_k\} | S_i^k(q) = S\}} x_{S_i^k}(t), \quad q = 1, \dots, N,$$

$$\langle S_q I_r \rangle(t) := \sum_{k=1}^{N-1} \sum_{\{i \in \{1, \dots, d_k\} | S_i^k(q) = S, S_i^k(r) = I\}} x_{S_i^k}(t), \quad q = 1, \dots, N, r = 1, \dots, N.$$

Tehát $\langle S_q \rangle$ jelöli azon állapotok valószínűségének összességét, melyekben a q -adik csúcs működő cég illetve $\langle S_q I_r \rangle$ annak a valószínűségét, hogy a q -adik csúcs működik, az r -edik csődben van a t időpontban. A (6.2) alapegyenlet-rendszerből kiindulva szeretnénk $\langle I_q \rangle$ -ra felírni egy differenciálegyenlet-rendszert:

14. Tétel. Tegyük fel, hogy teljesülnek a (6.2) egyenletek. Ekkor az $\langle I_q \rangle$ függvényekre teljesülnek az alábbi differenciálegyenletek:

$$\langle \dot{I}_q \rangle(t) = \lambda_q \langle S_q \rangle(t) + \sum_{p=1, p \neq q}^N \lambda_q \omega_{p,q} \langle S_q I_p \rangle(t), \quad q = 1, \dots, N. \quad (6.6)$$

BIZONYÍTÁS. Nyilván (6.2) miatt

$$\begin{aligned} \langle \dot{I}_q \rangle(t) &= \sum_{k=1}^N \sum_{\{i \in \{1, \dots, d_k\} | S_i^k(q)=I\}} \dot{x}_{S_i^k}(t) = \\ &= \sum_{k=1}^N \sum_{\{i \in \{1, \dots, d_k\} | S_i^k(q)=I\}} \left(\sum_{h \in H_{S_i^k}^-} f_{SI}(l_h, G, S_h^{k-1}) x_{S_h^{k-1}}(t) - \left(\sum_{j \in H_{S_i^k}^+} f_{SI}(l_j, G, S_i^k) \right) x_{S_i^k}(t) \right), \end{aligned} \quad (6.7)$$

ahol

$$\begin{aligned} f_{SI}(l_h, G, S_h^{k-1}) &= \lambda_{l_h} \left(1 + \sum_{\{p \in \{1, \dots, N\} | S_h^{k-1}(p)=I\}} \omega_{p, l_h} \right), \\ f_{SI}(l_j, G, S_i^k) &= \lambda_{l_j} \left(1 + \sum_{\{p \in \{1, \dots, N\} | S_i^k(p)=I\}} \omega_{p, l_j} \right). \end{aligned}$$

A jobb oldalon ki fog esni a tagok egy része, ugyanis ha tekintjük a (6.2) rendszer S^k állapothalmazhoz tartozó egyenletek jobboldalát összeadva, majd a negatív előjelű tagokat, azaz az $S^k \rightarrow S^{k+1}$ átmenetért felelős tagokat, akkor az összeg ezen tagjai megtalálhatóak lesznek a (6.2) rendszer S^{k+1} állapothalmazhoz tartozó egyenletek jobboldalában, ha ott szintén az $S^k \rightarrow S^{k+1}$ átmenetért felelős tagokat vesszük (ezek a pozitív előjelűek).

Nézzük meg ezt részletesebben. Az $S^k \rightarrow S^{k+1}$ átmenetért felelős tagok a (6.2) rendszer S^k állapothalmazhoz tartozó egyenleteinek jobboldalain rögzített k -ra:

$$- \sum_{\{i \in \{1, \dots, d_k\} | S_i^k(q)=I\}} \left(\left(\sum_{j \in H_{S_i^k}^+} \lambda_{l_j} \left(1 + \sum_{\{p \in \{1, \dots, N\} | S_i^k(p)=I\}} \omega_{p, l_j} \right) \right) x_{S_i^k}(t) \right), \quad (6.8)$$

míg az $S^k \rightarrow S^{k+1}$ átmenetért felelős tagok a (6.2) rendszer S^{k+1} állapothalmazhoz tartozó egyenleteinek jobboldalain:

$$\sum_{\{j \in \{1, \dots, d_{k+1}\} | S_j^{k+1}(q)=I\}} \left(\sum_{i \in H_{S_j^{k+1}}^-} \lambda_{l_i} \left(1 + \sum_{\{p \in \{1, \dots, N\} | S_i^k(p)=I\}} \omega_{p, l_i} \right) x_{S_i^k}(t) \right). \quad (6.9)$$

Mindkét esetben az $S^k \rightarrow S^{k+1}$ átmenet egy $S_i^k \rightarrow S_j^{k+1}$ közvetlen állapotváltozást jelent, ahol az $S_i^k \in S^k$, $S_j^{k+1} \in S^{k+1}$ általános jelöléseket alkalmazzuk a Markov-folyamat aktuális állapotaira.

Az első esetben (lásd (6.8)) olyan S_i^k állapotok szerepelnek az S^k halmazból, melyek q -adik eleme I és minden ilyen $i \in \{1, \dots, d_k\}$ esetén hozzárendeljük azokat az állapotokat S^{k+1} halmazból, melyek pontosan egy elembe térnek el S_i^k -től. Tehát vesszük azokat az S_i^k állapotokat, melyekre $S_i^k(q) = I$ és minden lehetséges módon elhelyezünk még egy I -t valamelyik S helyett az N csúcsú gráfon. Tehát $|\{i \in \{1, \dots, d_k\} \mid S_i^k(q) = I\}| = \binom{N-1}{k-1}$ és bármely rögzített S_i^k állapot esetén $|H_{S_i^k}^+| = N - k$, azaz a (6.8) összeg tagjainak száma: $\binom{N-1}{k-1} \cdot (N - k)$.

A második esetben (lásd (6.9)) az S^{k+1} halmaznak az összes olyan elemét tekintjük, melynek q -adik eleme I , majd vesszük az S^k halmazból azokat az állapotokat, melyek pontosan egy elembe térnek el ezek valamelyikétől. Tehát itt az S^{k+1} halmazból indulunk ki, vesszük azokat az elemeket, melyekre $S_j^{k+1}(q) = I$ és az egyik I helyére S -et írunk az N csúcsú gráfban. Vagyis $|\{j \in \{1, \dots, d_{k+1}\} \mid S_j^{k+1}(q) = I\}| = \binom{N-1}{k}$ és tetszőleges S_j^{k+1} állapot esetén $|H_{S_j^{k+1}}^-| = k + 1$, tehát a (6.9) összeg tagjainak száma: $\binom{N-1}{k} \cdot (k + 1)$.

Az első esetbeli S^k állapotok megtalálhatóak a második esetbeli S^k állapotok között. Azonban visszafelé ez nem igaz: veszek egy S_j^{k+1} állapotot, melyre $S_j^{k+1}(q) = I$, majd hozzárendelek egy S^k -beli elemet a $H_{S_j^{k+1}}^-$ halmazból, azaz olyat, hogy pontosan egy csúcs állapotában térjenek el: legyen ez az állapot S_i^k és a q -adik csúcsnál térjen el S_j^{k+1} állapottól. Ez azt jelenti, hogy nyilván $S_j^{k+1}(q) = I$ viszont $S_i^k(q) = S$, tehát ez az állapot nincs benne az első esetbeli S^k állapotok között, csak a második esetbeliben, és ilyenkor pont a q -adik csúcs állapotváltozása következik be. A kimaradó S^k állapotok indexeit az

$$\bigcup_{\{j \in \{1, \dots, d_{k+1}\} \mid S_j^{k+1}(q) = I\}} \{i \in H_{S_j^{k+1}}^- \mid S_i^k(q) = S\} \quad (6.10)$$

halmaz tartalmazza. Ellenőrzésképpen, a megmaradt tagok száma a (6.7) egyenlet jobboldalán az egyszerűsítések után a fenti leszámítások különbségéből:

$$\begin{aligned} \binom{N-1}{k} \cdot (k+1) - \binom{N-1}{k-1} \cdot (N-k) &= \frac{(N-1)!}{k!(N-1-k)!} (k+1) - \frac{(N-1)!}{(k-1)!(N-k-1)!} = \\ &= \frac{(N-1)!}{k!(N-1-k)!} = \binom{N-1}{k}, \end{aligned}$$

ami megegyezik a (6.10) halmaz elemszámával.

Ez azt jelenti, hogy a (6.7) egyenlet jobb oldalán az összes tag ki fog esni a (6.8) összegből, de a (6.9) összegben megmaradnak S^k azon elemei, melyeknek a q -adik csúcsa S és az $S^k \rightarrow S^{k+1}$ állapotváltozás során pont a q -adik csúcs fertőződik meg. Azaz az összevonások után a (6.7) egyenlet a következő alakot ölti:

$$\begin{aligned} \langle \dot{I}_q \rangle(t) &= \sum_{k=0}^{N-1} \sum_{\{i \in \{1, \dots, d_k\} | S_i^k(q)=S\}} \lambda_q (1 + \sum_{\{p \in \{1, \dots, N\} | S_i^k(p)=I\}} \omega_{p,q}) x_{S_i^k}(t) = \\ &= \sum_{k=0}^{N-1} \sum_{\{i \in \{1, \dots, d_k\} | S_i^k(q)=S\}} \left(\lambda_q x_{S_i^k}(t) + \sum_{\{p \in \{1, \dots, N\} | S_i^k(p)=I\}} \lambda_q \omega_{p,q} x_{S_i^k}(t) \right). \end{aligned}$$

Most rendszerezzük a tagokat együtthatók alapján. A λ_q együtthatóhoz tartozó tagok:

$$\sum_{k=0}^{N-1} \sum_{\{i \in \{1, \dots, d_k\} | S_i^k(q)=S\}} x_{S_i^k} = \langle S_q \rangle.$$

Rögzített p -re és q -ra ($p = 1, \dots, N$, $p \neq q$ és $q = 1, \dots, N$) a $\lambda_q \omega_{p,q}$ együtthatóhoz tartozó tagok pont azok az elemei az állapot térnek melyeknek q -adik eleme S és p -edik eleme I , azaz:

$$\sum_{k=1}^{N-1} \sum_{\{i \in \{1, \dots, d_k\} | S_i^k(q)=S, S_i^k(p)=I\}} x_{S_i^k} = \langle S_q I_p \rangle.$$

Tehát

$$\langle \dot{I}_q \rangle(t) = \lambda_q \langle S_q \rangle(t) + \sum_{p=1, p \neq q}^N \lambda_q \omega_{p,q} \langle S_q I_p \rangle(t), \quad q = 1, \dots, N.$$

□

Megjegyezzük, hogyha a kapcsolati háló homogén és a spontán csődbe jutási ráták is azonosak, azaz $\omega_{p,q} \equiv \omega$, $\lambda_q \equiv \lambda$, $p, q = 1, \dots, N$ valamely ω és λ mellett, akkor a (6.6) rendszerből levezethető egy pontos átlagoláson alapuló egyenlet az átlagos betegszámra:

$$\langle \dot{I} \rangle(t) = \lambda \langle S \rangle(t) + (N-1) \lambda \omega \langle SI \rangle(t),$$

amennyiben a dinamika SI típusú és a (6.1) függvénycsaláddal adjuk meg az átmeneti rátákat.

A (6.6) egyenletrendszer alapján megadunk egy közelítő differenciálegyenlet-rendszert is, melynek megoldásai az $\langle I_q \rangle$ valószínűségeket közelítik.

15. Következmény. Felhasználva az $\langle S_q \rangle = 1 - \langle I_q \rangle$ összefüggést és az $\langle S_q I_r \rangle \approx \langle S_q \rangle \cdot \langle I_r \rangle$ közelítést az egyenletek lezárására, a (6.6) egyenlet egy közelítését kapjuk. A közelítő egyenletek a következő alakúak:

$$\dot{x}_q(t) = \lambda_q(1 - x_q(t)) \left(1 + \sum_{p=1, p \neq q}^N \omega_{p,q} x_p(t) \right), \quad q = 1, \dots, N, \quad (6.11)$$

ahol az x_q függvények jelölik a közelítő egyenlet megoldásait.

BIZONYÍTÁS. Mivel $\langle I_q \rangle \approx x_q$, $\langle S_q \rangle \approx 1 - x_q$ és $\langle S_q I_p \rangle \approx \langle S_q \rangle \cdot \langle I_p \rangle = (1 - x_q)x_p$, ezért a közelítő differenciálegyenlet-rendszer:

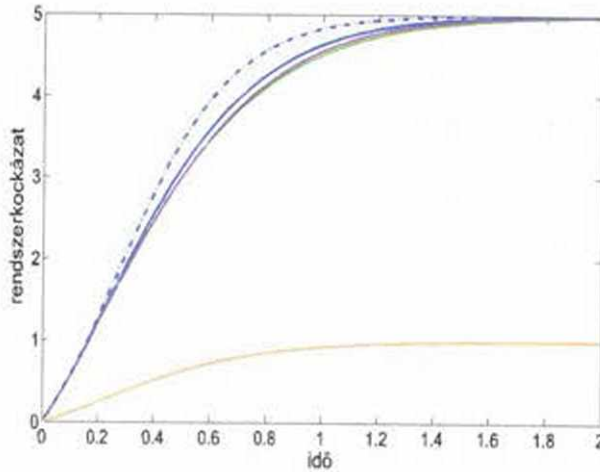
$$\dot{x}_q(t) = \lambda_q(1 - x_q(t)) + \sum_{p=1, p \neq q}^N \lambda_q \omega_{p,q} (1 - x_q(t)) x_p(t).$$

□

A (6.11) differenciálegyenlet-rendszer megoldása után a rendszerkockázatot könnyedén számolhatjuk a

$$K(t) = \sum_{q=1}^N x_q(t)$$

képlet segítségével, hiszen x_q közelítőleg meghatározza a q -adik csúcs csödjének a valószínűségét. A (6.11) modell előnye, hogy tetszőlegesen nagy N esetén is gyorsan megoldható numerikusan.



6.2. ábra. Rendszerkockázat az idő függvényében teljes (kék), csillag (zöld) és körgráfon (magenta) a (6.2) alapegyenletekkel (egyenes) és a (6.11) modellel (szaggatott görbe), $N = 5$, $\lambda_q = 1$, éleken az összszúly: $2\omega N = 2 \cdot 2 \cdot 5$. Illetve egy cég esetén a csőd valószínűségének a görbéje (barna) teljes gráfon. 0 kezdeti csödből indítva a folyamatot.

A 6.2. ábrán összevetjük a (6.2) és a (6.11) modellek eredményeit egyforma spontán csődbe menési ráták mellett különböző gráftípusokon: teljes, csillag és körgráfon rögzített összsúly mellett, tehát ha $\sum_{p=1}^N \sum_{q=1}^N \omega_{p,q}$ adott. Látható, hogy a rendszerkockázati görbék kvalitatíve nagyon hasonlóak, így a (6.11) modell alkalmas a (6.2) rendszer helyettesítésére nagy N esetén. Az ábrán még azt érdemes észrevenni, hogy a (6.11) rendszer ugyanolyan eredményt ad kör- és teljes gráf esetén (a görbék teljesen egymáson vannak), míg a (6.2) modellel különböző rendszerkockázati görbéket kapunk. Ennek magyarázata lejjebb található. A következő részben a (6.11) közelítő rendszer viselkedését vizsgáljuk meg különböző struktúrájú gráfok esetén.

6.4.2. A modell vizsgálata

Ebben a fejezetben egyrészt explicit képletet adunk a rendszerkockázat kiszámítására a (6.11) rendszer alapján néhány egyszerű esetben, másrészt arra keressük a választ, hogy csökkenthető-e a rendszerkockázat a gráf struktúrájának megfelelő megválasztásával, illetve speciális struktúrájú gráfok esetén hogyan alakul a várhatóan csődbe lévő cégek száma az idő függvényében.

A közelítő egyenletrendszer analízise egyszerű esetekben

Először néhány egyszerű megállapítást teszünk, ugyanis ha a cégek sérülékenysége azonos, akkor a (6.11) differenciálegyenlet-rendszer leegyszerűsödik, így megoldható elemi módszerek segítségével. Két lemmát fogalmazunk meg - melyeket egyszerűségükből adódóan bizonyítás nélkül közlünk - az első homogén, míg a másodikat heterogén kapcsolati háló esetén.

16. Lemma. *Ha $\lambda_q \equiv \lambda$, $\omega_{p,q} \equiv \omega$, $p, q = 1, \dots, N$, valamely λ és ω mellett és a kezdeti feltétel a szokásos, azaz $x_q(0) = 0$, $q = 1, \dots, N$, akkor az összes cég helyzete ugyanúgy fog alakulni, azaz ilyenkor $x_q(t) = x(t)$ teljesülni fog valamely $x(t)$ függvény esetén. Felhasználva a feltételeket a (6.11) differenciálegyenlet-rendszerből egy szétválasztható változójú differenciálegyenlethez jutunk:*

$$\dot{x}(t) = \lambda(1 - x(t)) \left(1 + (N - 1)\omega x(t) \right).$$

Melynek megoldása:

$$x(t) = \frac{\lambda(1 + (N - 1)\omega)}{K e^{t\lambda(1 + (N - 1)\omega)} - (N - 1)\lambda\omega} + 1.$$

Az $x(0) = 0$ kezdeti feltétel mellett $K = -\lambda$, így a megoldás:

$$x(t) = \frac{\lambda(1 + (N - 1)\omega)}{-\lambda e^{t\lambda(1 + (N - 1)\omega)} - (N - 1)\lambda\omega} + 1.$$

Ekkor a rendszerkockázat képlete: $K(t) = Nx(t) = N \cdot \left(-\frac{1+(N-1)\omega}{e^{t\lambda(1+(N-1)\omega)} + (N-1)\omega} + 1 \right)$.

17. Lemma. Ha $\lambda_q \equiv \lambda$, $q = 1, \dots, N$, valamely λ mellett és a kezdeti feltétel $x_q(0) = 0$, $q = 1, \dots, N$, akkor a (6.11) rendszerből a következő N változós differenciálegyenlet-rendszerhez jutunk:

$$\dot{x}_q(t) = \lambda(1 - x_q(t)) \left(1 + \sum_{p=1, p \neq q}^n \omega_{p,q} x_p(t) \right), \quad q = 1, \dots, N. \quad (6.12)$$

Ha a $\sum_{p=1, p \neq q}^n \omega_{p,q}$ érték, azaz az egy csúcsba befutó súlyok összege független q -tól, tehát $\sum_{p=1, p \neq q}^n \omega_{p,q} = B$, minden $q = 1, \dots, N$ esetén, akkor $x_q(t) = y(t)$, $q = 1, \dots, N$ teljesülni fog a (6.12) rendszer megoldásaira valamely olyan $y(t)$ függvény esetén, melyre

$$\dot{y}(t) = \lambda(1 - y(t))(1 + By(t)),$$

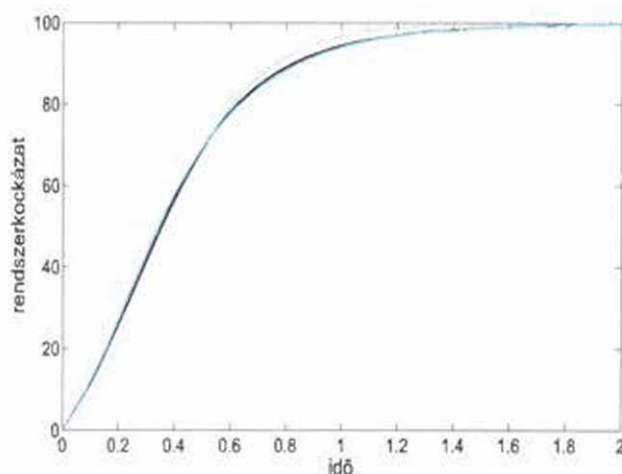
vagyis minden $x_q(t)$ függvényre ugyanaz a megoldást kapjuk: $y(t) = -\frac{1+B}{e^{t\lambda(1+B)} + B} + 1$. Ezek alapján a rendszerkockázat is kiszámítható:

$$K(t) = Ny(t) = N \cdot \left(-\frac{1+B}{e^{t\lambda(1+B)} + B} + 1 \right). \quad (6.13)$$

Ezzel magyarázható, hogy ha az összsúlyt lerögzítjük, akkor például körgráfon és teljes gráfon ugyanazt az eredményt kapjuk, ha az idő függvényében ábrázoljuk a rendszerkockázatot a (6.11) rendszer alapján.

Speciális struktúrájú gráfok esete

Ebben részben a rendszerkockázati görbe változását követjük nyomon a gráfstruktúra függvényében. Nyilván, ahogy az éleken található összsúlyt növeljük, azaz egyre szorosabb kapcsolatot teszünk fel a cégek között, úgy a rendszerkockázat egyre jobban növekszik, tehát adott idő alatt várhatóan egyre több cég kerül csődbe. Ugyanez mondható el a cégek sérülékenységének növelése esetén. Ezért állandó összsúly és rögzített spontán csődbe menési ráták mellett fogjuk tekinteni a különböző gráftípusokon a rendszerkockázat időbeli lefolyását. A kérdés az, hogy hogyan érdemes beállítani a kapcsolati hálót, hogy minél kisebb legyen a várhatóan csődbe lévő cégek száma. Speciális struktúrájú irányítatlan gráfokon, azaz teljes, kör- illetve csillag gráfon vizsgáljuk a folyamatot, illetve véletlen gráfokon: reguláris és bimodális véletlen gráfon illetve Barabási-Albert gráfon. Az összsúly legyen: $\sum_{p=1}^N \sum_{q=1}^N \omega_{p,q} = 2\omega N$ és $\lambda_q = 1$, $q = 1, \dots, N$, a gráfok eloszlásának megfelelően elosztva az egyes éleken, példaként teljes gráf



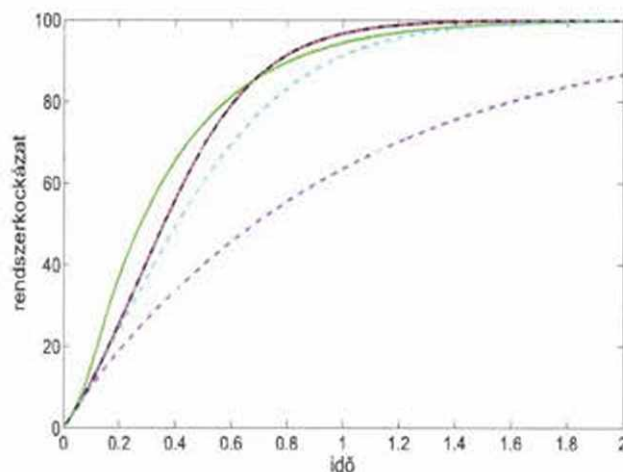
6.3. ábra. Rendszerkockázat az idő függvényében bimodális (fekete) ($n_1 = 2$, $n_2 = 6$) és Barabási-Albert (cián) gráfon, emellett teljes (kék szaggatott görbe) és csillag (zöld szaggatott görbe) gráfon, $N = 100$, $\lambda_q = 1$, $q = 1, \dots, N$, $2\omega N = 2 \cdot 2 \cdot 100$, 0 kezdeti csődből indítva.

esetén egy élen $\frac{2\omega N}{N(N-1)} = \frac{2\omega}{N-1}$ súly lesz (oda-vissza is), míg körgráf esetén $\frac{2\omega N}{2N} = \omega$ és csillag gráf esetén $\frac{2\omega N}{2(N-1)} = \frac{\omega N}{N-1}$.

A numerikus eredményekből látható, hogy az idő függvényében ábrázolt rendszerkockázati görbe nagyon hasonló a vizsgált gráfokon, így a kaotikusság elkerülése végett a 6.3. ábrán csak néhány gráftípus rendszerkockázati görbét adtuk meg. A vizsgálat alapján néhány apró észrevételt lehet tenni: a homogén súlyozású esetekben, vagyis a teljes és a reguláris véletlen gráfon a rendszerkockázat előáll a (6.13) képlet alapján, ahol $B = 2\omega$, hiszen ezekben az esetekben az egy csúcsból kiinduló élsúlyok összege állandó. Sőt, még a körgráf rendszerkockázata is egybeesik az előbbiekével, mivel itt is az egy csúcsba befutó súlyok összege független a csúcstól és ez az érték megegyezik mindhárom gráf esetén. Tehát ez a három grafikon egybeesik. Ezen görbéhez képest elemezzük a többi gráfhhoz tartozó rendszerkockázati görbe helyzetét. Kezdetben a heterogén súlyozású gráfok a (6.13) görbe felett vannak, minél heterogénebb (csillag, majd Barabási-Albert, végül a bimodális gráf), annál inkább felette, majd a (6.13) görbe alámennek (ugyanilyen sorrendben).

A rendszerkockázati görbék hasonlósága arra enged következtetni, hogy irányítatlan gráfok esetén, ha az összsúly fix, akkor a kapcsolati háló speciális beállításával se lehet a rendszerkockázatot látványosan lecsökkenteni. Ez motiválja, hogy megvizsgáljuk ugyanezt a kérdés irányított, speciális struktúrájú gráfok esetén is. Csak néhány jól átlátható esettel foglalkoz-

tunk: körgráf esetén azt vizsgáltuk, amikor a cégek között körbeírányítva mennek az élek, azaz van benne pontosan egy irányított kör. Csillag gráfnál két esetet néztünk meg: az egyik esetben a gráfon minden él egy központi csúcs felé van befelé irányítva (a továbbiakban be-csillag gráfként hivatkozunk rá), a másik esetben pedig ha kifelé irányítjuk az éleket a központi csúcsból (a továbbiakban ki-csillag gráfként említjük). Természetesen itt is feltettük, hogy az összecsúszási állandó és a sérülékenységi rátákat megint fixáltuk.



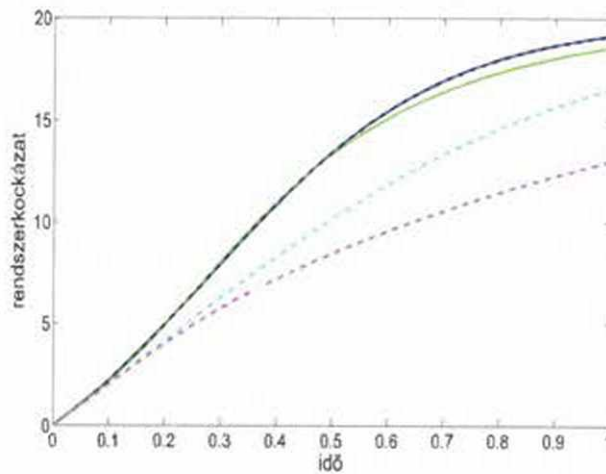
6.4. ábra. Rendszerkockázat az idő függvényében a (6.13) egyenletből kapott esetben (piros görbe), csillag gráfon (zöld), irányított kör- (szaggatott fekete) és ki-csillag (szaggatott cián), be-csillag (szaggatott lila) gráfon, $N = 100$, $\lambda_q = 1$, $q = 1, \dots, N$, $2\omega N = 2 \cdot 2 \cdot 100$, 0 kezdeti csődből indítva.

A 6.4. ábrán az látható, hogy az irányított körgráfon kapott megoldás a (6.13) görbét adja $B = 2\omega$ mellett, hiszen itt is az egy csúcsba befutó súlyok összege csúcsfüggetlen, azaz független a csúcsától. Az aszimmetria csökkenti a rendszerkockázatot a csillag gráf esetében: ha a középponti csúcs csődje van csak hatással a többi csúcs csődjére, az már érzékelhető csökkenést okoz (ki-csillag), ugyanis csak a középponti csúcs csődje esetén kezd el jelentősen összeomlani a rendszer, de ennek a csúcsnak a csődjét csak a saját spontán csődbe jutási rátája befolyásolja. Ha csak a középponti cég csődje van hatással a többi periférián található cég, akkor lényegében a külső $N - 1$ cég teljesen független egymástól, csődjüket más cégek összeomlása nem befolyásolja, így ez a gráf struktúra csökkenti a legjobban a rendszerkockázatot (be-csillag).

Az utolsó vizsgált esetben már nem csak a cégek egymáshoz való viszonya lehet különböző, hanem a különböző cégek sérülékenységi rátája is. Tehát ha egy cég spontán csődbe

menési rátája kicsi, az azt jelenti, hogy a cég erős, nehezen kerül csődbe, míg nagy λ_q értékkel szerepelnek a gyenge cégek.

Itt is néhány pár szélsőséges esetet hasonlítottunk össze: irányítatlan teljes, csillag, illetve irányított kör-, ki-csillag és be-csillag gráfon. A cégek közötti erőviszony: az első cég spontán csődbe menési rátája fele akkora mint a többi cégnek, tehát az első cég a legerősebb. Ki- és be-csillag gráf esetén az erős cég elhelyezkedése is fontos: ez a cég van középen. Az összevetésbe néhány irányítatlan gráfot is belevettünk azért, hogy pontosabb kép alakuljon ki az irányított gráfok esetéről. A 6.5. ábrán látható hogy a teljes, illetve az irányított körgráf majdnem ugyanúgy szerepel és ők a legrosszabbak rendszerkockázat szempontjából. A ki-csillag gráf esetén a központi erős cég csődje van hatással a periférián lévő gyengébb cégekre, ez tovább csökkenti a rendszerkockázatot az előző esetbeli ki-csillag rendszerkockázatához képest, mert az erős cég működése még stabilabb, mint egyforma λ_q értékek esetén. Ezzel szemben a be-csillag gráfon a központi erős cégre vannak hatással a gyenge cégek, amik gyorsabban csődbe kerülnek, így ez növeli kicsit a rendszerkockázatot az előző esetbeli be-csillag gráfon mérthez képest, tehát a ki-csillag és be-csillag gráfok rendszerkockázati görbéi egy erős cég esetén közelednek egymáshoz, az egyforma λ_q értékek megadó esetéhez képest, de még így is a be-csillag gráf a legoptimálisabb struktúrájú gráf a kockázat szempontjából.



6.5. ábra. Rendszerkockázat az idő függvényében teljes (kék), csillag (zöld), irányított kör- (szaggatott fekete) és be-csillag gráfon (szaggatott magenta), ki-csillag (szaggatott cián), $N = 20$, $\lambda_1 = 0.5$, $\lambda_q = 1$, $q \neq 1$, $q = 1, \dots, N$, $2\omega N = 2 \cdot 2 \cdot 20$, 0 kezdeti csődből indítva.

6.5. Összefoglalás

A fejezet során két témakörrel foglalkoztunk: egyrészt egy hálózati modellt építettünk fel egy pénzügyi probléma megoldására, másrészt pedig az új modellt alkalmaztuk egy másik fontos pénzügyi kérdés vizsgálatára.

A 6.2. szakaszban bemutattunk egy valószínűségszámításon alapuló eljárást, ami az egyes cégek adott időpontbeli csődbe jutási esélyeiből megadja az együttes csődvalószínűségeket ugyanabban az időpontban. A következő szakaszban ajánlottunk egy hálózati modellt ugyanazon probléma megoldására. A modellünk előnye, hogy - ellentétben az eredeti, pénzügyi piacokon alkalmazott eljárással - nem csak a kívánt együttes bedőlési valószínűségeket adja meg, hanem betekintést enged a teljes folyamat működésébe is. Pontosabban a hálózatos módszer során felírt lineáris differenciálegyenlet-rendszerből kiszámítható a kapcsolódó 2^N méretű Markov-folyamat összes állapotának valószínűsége minden időpontban. Ezért sokkal több információt megtudhatunk a cégek csődbe jutásának lezajlásáról, például a rendszerkockázati görbét is könnyen kiszámíthatjuk. Ezenkívül a kínált eljárás lehetőséget teremt heterogén kapcsolati háló vizsgálatára is. A módszer hátránya, hogy csak kis számú cég esetén kivitelezhető.

A fejezet további részében kidolgoztunk egy nagy N esetén is numerikusan jól kezelhető közelítő technikát a rendszerkockázat meghatározására. Itt egy új problémával kezdtünk el foglalkozni, miszerint a gráfstruktúra hogyan befolyásolja a csődben lévő cégek számának várható értékét. Megvizsgáltuk különböző típusú gráfokon a rendszerkockázat mértékét, és arra a következtetésre jutottunk, hogy rögzített összelsúly mellett irányított gráfokon van esély a kockázat csökkentésére és kiemelkedő eredményt a be-csillag gráfon lehet elérni. A modell egy kiterjesztése lehet, ha megengedünk negatív súlyokat is a gráfban. Ez azt jelentené, hogy ha negatív súly van két cég között, akkor az egyik cég csődje a másik csődjének esélyét csökkenti, azaz a cég ilyenkor megerősödik, ezt úgy lehet elképzelni, hogy például megkapja a csődbe került cég befektetőit. Tehát negatív ω akkor van két cég között, ha a két cég konkurenciái egymásnak. Ezzel szemben, ha az élsúly pozitív, akkor a két cég egymás partnere, tehát az egyik cég csődje, a másik cég csődjének esélyét növeli. Ezek a gondolatok egy további lehetséges kutatási irányt adnak meg.

Összefoglalás

Az értekezés során különböző hálózatokon zajló folyamatokat vizsgáltunk *SIS* és *SI* típusú dinamika mellett. A matematikai modell, amivel leírjuk ezeket a folyamatokat egy folytonos paraméterű Markov-lánc, 2^N elemszámú állapottérrel egy N csúcsú hálózat esetén. A folyamathoz tartozó alapegyenlet-rendszer mérete 2^N , ezért numerikusan kezelhetetlen N tipikus értékeire. Ezért fontos minél kisebb dimenziós modellek kidolgozása a hálózatokon zajló folyamatok vizsgálatára.

Doktori kutatásom egyik célkitűzése az alapegyenletek számának jelentős csökkentése különböző struktúrájú gráfok esetén, olyan eljárásokkal, amik megőrzik a folyamat dinamikájának legfőbb jellemvonásait. Ezen témakörben egy hatékony módszert fejlesztettünk ki tetszőleges fokszámoszlású konfigurációs modellel generált gráfon zajló betegségterjedés modellezésére *SIS* dinamika esetén. A fő kihívást, adott betegszám mellett, az *SI* élek átlagos számának meghatározása jelentette, ugyanis ezeken az éleken keresztül terjed tovább a betegség, így ezen mennyiség ismeretével a folyamat működésébe nyerhetünk betekintést. Az alapegyenletek számának eredményes redukcióját értük el körgráfon is, arra a megfigyelésre alapozva a modellt, hogy a betegség frontszerűen halad végig a körgráfon. Mindkét munka során a redukált egyenletrendszerek mérete N nagyságrendű, így numerikusan is jól kezelhető technikákat kaptunk és az átlagos betegszám az idő függvényében megadható.

A dolgozat következő részében ismert, alacsony dimenziós rendszerek elemzésével foglalkoztunk, fő tekintettel az párok szintjén felírt kompakt egyenletrendszerre, ami egy jól használható modell, de analitikus vizsgálata még nem teljes.

Az utolsó munkánkban arra kerestük a választ, hogy a pénzügyi piacokon megjelenő problémákat, mint a hitelbiztosítások árazása, illetve a rendszerkockázat mérése, hogyan lehet leírni hálózati modellek segítségével. A kutatások során kidolgozott modell sokkal általánosabb vizsgálatot tesz lehetővé a pénzügyi életben használatos eljárásnál.

Summary

In the PhD dissertation epidemic processes running on networks were examined for *SIS* and *SI* dynamics. The mathematical model describing these processes is a continuous time Markov chain with an extremely large state space, leading to the master equations consisting of 2^N equations, where N is the number of nodes in the network. Solving these even numerically is impossible for the typical values of N , thus it is important to derive low-dimensional ODE approximations for modeling these processes.

The first aim of the PhD dissertation is to reduce the size of the master equations by introducing new methods, that can capture the main features of the process. In this topic, a viable approach is presented for graphs with arbitrary degree distributions built according to the configuration model, when *SIS* dynamics is considered. Here, the main challenge is to compute the average number of *SI* edges, when the number of infected nodes is given. Significant reduction of the size of the master equations is achieved on cycle graph, based on the observation that the epidemic spreads along the cycle graph as a front. In every method, the size of the new approximating systems is of order N , thus they are numerically solvable.

In the next thesis known, low-dimensional ODE approximations are studied and a compact pairwise approximation is examined partially: uniqueness of the endemic steady state is proved. Since this system is easily solvable numerically, but its analytical investigation is not completed, it is worthwhile to continue the investigation of the analysis of the system.

In the last part of the work, a network approach is presented to handle a financial problem, namely, the process of joint bankruptcy of N firms. The most important question of the thesis is how we can connect the network theory with the existing financial approaches. It turned out that the network approach enables more general examinations of the financial processes.

Irodalomjegyzék

Az értekezés alapjául szolgáló saját publikációk

- [1] N. Nagy, I.Z. Kiss, P.L. Simon, Approximate master equation for dynamical processes on graphs, *Math. Model. Nat. Phenom.*, **9** 02 (2014), 43-57.
- [2] N. Nagy, P.L. Simon. Monte-Carlo simulation and analytic approximation of epidemic processes on large networks, *Cent. Eur. J. Math.*, **11** 4 (2013), 800-815 .
- [3] N. Nagy, P.L. Simon, Detailed analytic study of the compact pairwise *SIS* epidemic model. (előkészületben)
- [4] N. Nagy, P.L. Simon, Zs. Bihary, Network process approach for BDS pricing. (előkészületben)

További hivatkozások

- [5] R.M. Anderson, R.M. May, Infectious Diseases of Humans Dynamics and Control, Oxford University Press, (1992).
- [6] F. Ball, Stochastic and deterministic models for SIS epidemics among a population partitioned into households, *Math. Biosci.*, **156** (1999), 41-68.
- [7] A.-L. Barabási, R. Albert, Emergence of scaling in random networks, *Science*, **286** (5439) (1999), 509-512.
- [8] A. Barrat, M. Barthélemy, A. Vespignani, Dynamical Processes on Complex Networks, Cambridge University Press, Cambridge, (2008).

- [9] B. Bollobás, A probabilistic proof of an asymptotic formula for the number of labelled regular graphs, *Europ. J. Combin.* **1** 4 (1980), 311-316.
- [10] B. Bollobás, R. Kozma, D. Miklós, Handbook of large-scale random networks, János Bolyai Math. Soc. and Springer-Verlag, Budapest, Berlin, (2009).
- [11] B. Bollobás, Random graphs, Cambridge University Press, Cambridge, (2001).
- [12] L. Danon, A.P. Ford, T. House, C.P. Jewell, M.J. Keeling, G.O. Roberts, J.V. Ross, M.C. Vernon, Networks and the Epidemiology of Infectious Disease, *Interdisciplinary Perspectives on Infectious Diseases* 2011:284909 special issue "Network Perspectives on Infectious Disease Dynamics", (2011).
- [13] O. Diekmann, J. A. P. Heesterbeek, Mathematical epidemiology of infectious diseases: model building, analysis and interpretation, John Wiley and Sons, (2000).
- [14] X. C. Fu, M. Small, G. R. Chen, Propagation dynamics on complex networks: Models, methods and stability analysis, John, Wiley and Sons, (2014).
- [15] G. Ghoshal, L. M. Sander, I. M. Sokolov, SIS epidemics with household structure: the self-consistent field method, *Math. Biosci.* **190** 1 (2004), 71-85.
- [16] T. House, M. J. Keeling, Insights from unifying modern approximations to infections on networks, *J. Roy. Soc. Interface* **8** (2011), 67-73.
- [17] T. House, M.J. Keeling, The impact of contact tracing in clustered populations, *PLoS Comput. Biol.* **6** (2010), e1000721.
- [18] M.J. Keeling, K.T.D. Eames, Networks and epidemic models, *J. Roy. Soc. Interface* **2** (2005), 295-307.
- [19] D. X. Li, On Default Correlation: A Copula Function Approach, *Journal of Fixed Income* **9** 4 (2000), 43-54.
- [20] I. Nåsell, The quasi-stationary distribution of the closed endemic SIS model, *Adv. Appl. Probab.* **28** (1996), 895 – 932.
- [21] M. E. J. Newman, A.-L. Barabási, D.J. Watts, The structure and dynamics of networks, Princeton University Press, Princeton, NJ, (2006).

- [22] M. E. J. Newman, *Networks: An Introduction*, Oxford, (2010).
- [23] R. Pastor-Satorras, A. Vespignani, Epidemic spreading in scale-free networks, *Phys. Rev. Lett.* **86** (2001), 3200 – 3203.
- [24] M. Roy, M. Pascual, On representing network heterogeneities in the incidence rate of simple epidemic models, *Ecol. Complexity*, **3** 1 (2006), 80-90.
- [25] K. J. Sharkey, Deterministic epidemic models on contract networks: Correlations and unbiological terms, *Theor. Popul. Biol.* **79** 4 (2011), 115-29.
- [26] P.L. Simon, M. Taylor & I.Z. Kiss, Exact epidemic models on graphs using graph-automorphism driven lumping, *J. Math. Biol.* **62** (2010), 479-508.
- [27] D. J. Watts, S. Strogatz, Collective dynamics of 'small-world' networks, *Nature* **393** (6684) (1998), 440-442.

